OPTIMAL MISSING PIXEL ESTIMATION ALGORITHMS FOR LARGE DETECTOR ARRAYS

FINAL PROGRESS REPORT

HAIRONG QI, WESLEY E. SNYDER, WILLIAM SANDER, GRIFF BILBRO

JULY 15, 1999

U.S. ARMY RESEARCH OFFICE

DAAH04-93-D-0003

CENTER FOR ADVANCED COMPUTING AND COMMUNICATION

NORTH CAROLINA STATE UNIVERSITY

RALEIGH, NC 27695-7914

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	07/15/99	Final Progress

4. TITLE AND SUBTITLE

Optimal Missing Pixel Estimation Algorithms for Large Detector Arrays

5. FUNDING NUMBERS

DAAH04-93-D-0003

6. AUTHOR(S)

Hairong Qi, Wesley Snyder, William Sander, Griff Bilbro

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

Center for Advanced Computing and Communication
Box 7914
North Carolina State University
Raleigh, NC 27695-7914

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 35741.4-EL-SR

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of optimal algorithm design is to develop a cost-effective high-resolution, large-area, digital imaging system, where optimal restoration methods are used to compensate hardware defects, so to reduce over-all system cost. Five modules are implemented in this imaging system: *image acquisition*, *image display*, *image enhancement*, *image correction*, and *image restoration*. The focus is on image correction and image restoration.

Besides extending the conventional polynomial approximation method to correct both radiometric and geometric distortions, we demonstrate the successful use of the thin-plate spline (TPS) interpolation method for geometric correction since TPS achieves *exact* mapping.

Optimal missing data estimation algorithms including deblurring and denoising are designed to restore images captured from large CCD sensor arrays using butting technique, where 1 to 2 columns of data are missed at the butting edge. We developed the *consistency method with separable deblurring*, which can deblur the original image and at the same time estimate the missing column(s) *exactly*, under the condition that no noise is inserted, and the separable blur kernel is exactly known. We also modified the maximum *a-posteriori* probability (MAP) estimate with the optimization problem solved by *mean field annealing* (MFA) to fit into this missing data estimation application. It shows more tolerance to perturbations due to noise and inaccurate blur kernel estimation.

14. SUBJECT TERMS

image processing, large detector arrays, image correction, image restoration, missing data estimation

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENT

LIST OF FIGURES

1. Problem Statement

The aim of this research work is to build and test a *large area, high resolution, digital* imaging system, which can be used in the Army and in medical care. Digital images are obtained from CCD sensor arrays. CCD imaging has many advantages over traditional film photography, such as its linearity, larger dynamic range and higher sensitivity. Despite all of these advantages, however, one of its major shortcomings is the low spatial resolution. "Butting" technology is an economic solution for this problem, where one or more CCD subarrays are butted together to achieve large area imaging (Fig. 1).

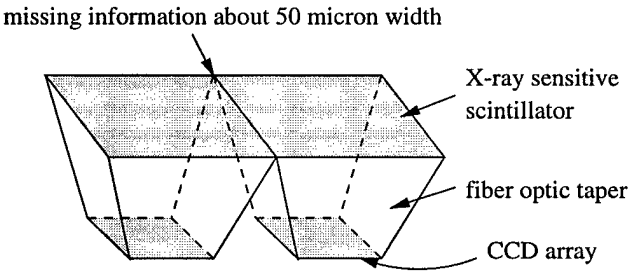


Figure 1. Demonstration of two scintillator/fiber/CCDs butted together.

Problem with butting is that no matter how close the two subarrays are, there will always be a gap on the order of one to two pixels which is about 50 micron wide. Therefore, optimal estimation algorithms are required to restore the missing data while at the same time recover the image from other defections, such as blur, noise, radiometric and geometric distortions.

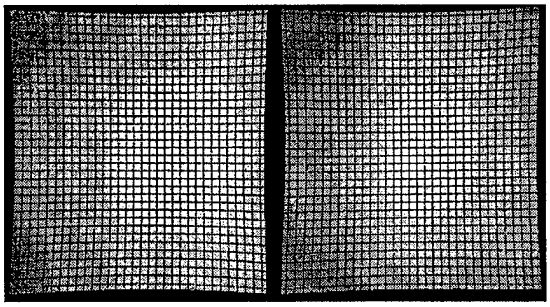


Figure 2. An original captured image from our imaging system.

Fig. 2 shows an original image captured from our imaging system. Severe vignetting-type radiometric distortion and pincushion-type geometric distortion can be observed. The problems we need to solve include:

- imaging system assembly;
- large area image display;
- radiometric correction;
- geometric correction; and
- missing data estimation along with deblurring and denoising.

2. Results Summary

Five modules are designed and implemented in this system: image acquisition, image display, image enhancement, image correction, and image restoration (Fig. 3).

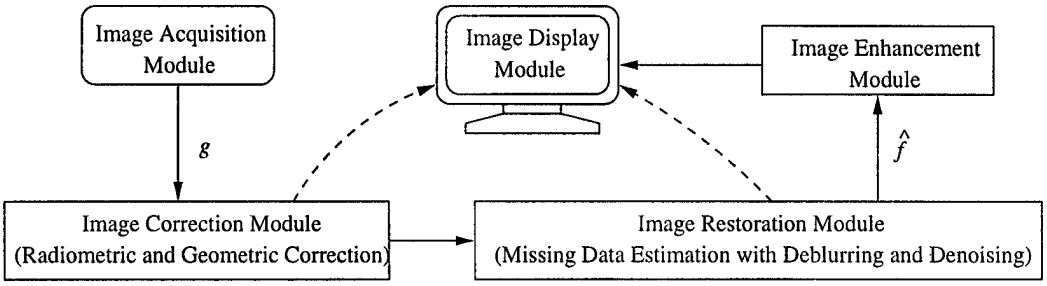


Figure 3. System module achievements.

The image acquisition module involves a 1 x 2 CCD/scintillator/fiber optic taper combinations, the CCD driver assembly, the A/D converter, and the control circuits.

In our system, the spatial resolution of a complete image is roughly 1152 x 2484, involving a 1 x 2 1152 x 1242 arrays. Since most of the current commercial monitors can only display image with dimensions 1024 x 1024 or smaller, how to efficiently and conveniently display larger images on smaller display planes is a problem which needs to be solved.

The image correction module consists of radiometric and geometric corrections. Both radiometric and geometric distortions are caused by the distorted shape of the fiber optic tapers. We use *polynomial approximation* to model a series of concentric ellipses formed by *vignetting-type* radiometric distortion. We use both *polynomial approximation* and *thin-plate spline interpolation* to correct the geometric distortion (mainly pincushion distortion). We show that thin-plate spline interpolation performs better than polynomial interpolation since it maps all the control points to their correspondence *exactly.*

The image restoration module is used to estimate missing data along with deblurring and denoising. We use two approaches to estimate missing data in a blurred, noisy image: the "consistency method with separable deblurring", and the "maximum *a-posteriori* probability (MAP) estimate with mean field annealing (MFA)". Both methods try to convert an ill-posed image restoration problem to a well-posed one by relaxation theory. The consistency method is able to recover missing data *exactly* from a blurred image based on a few assumptions. However, this exact restoration ability is limited by these assumptions, and the solution is unstable for large noise and inaccurate estimation of the point spread function (PSF). The MFA approach solves the MAP estimate by global optimization technique. It restores the missing data not so exact as the consistency method, but it is more stable to large noise and inaccurate estimation of the PSF.

Experiments are carried out on both synthetic data and real data. The experimental results show that the imaging system we developed exhibits good performance, and the algorithms high efficiency.

2.1 Image Acquisition and Display

Image acquisition is the first step in any imaging systems. For CCD imagers, the image acquisition module functions as a combination of photo-detector, amplifier, sampler, and quantizer. The image display module

reformulates the stream of sampled and quantized digital data to a 2-D/3-D matrix, and paints it on a 2-D image plane/computer monitor. The whole process is illustrated in Fig. 4.

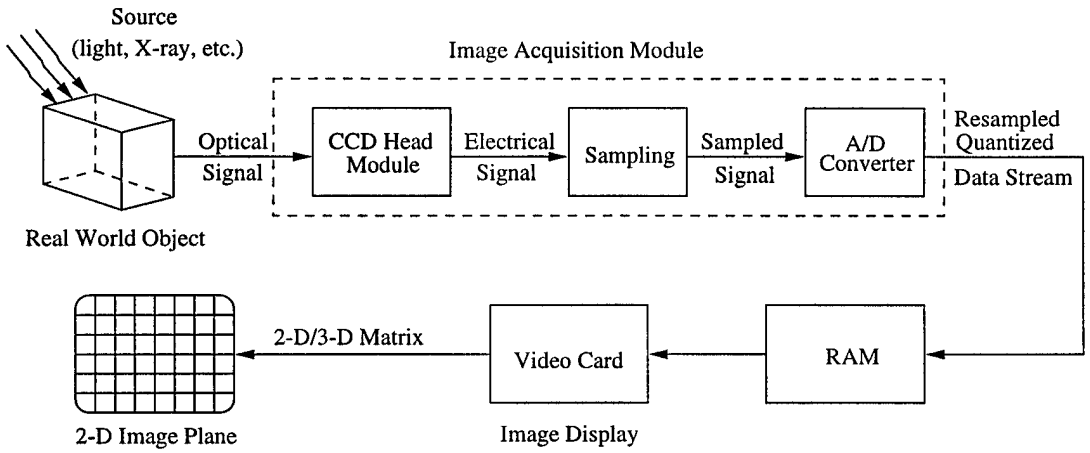


Figure 4. Image acquisition and display.

With a 1MHz readout rate, 1M sample rate, 1152 vertical pixels and 1242 horizontal pixels (equating to 1,430,784 total pixels in an array), the performance of our data acquisition module can be summarized as follows: 1.43s minimum is required to readout a complete frame; 1.24ms minimum is required to readout a complete line; and 1μs is required to readout 1 pixel of information. The detail design of image acquisition module is described in Appendix 7.1.

In image display module, we use a secondary display board combined with the host video card to achieve the so-called "dual-screen mode loupe" display. This technique can display complete images on host monitor, and segments of images on the secondary monitor for in-depth analysis, such as zooming, autoscaling, contrast stretching, pseudo-color enhancement, and missing data estimation. The design idea is explained in Appendix 7.2.

2.2 Image Correction

Based on the analysis of sources of radiometric and geometric distortions (Appendix 7.3), we implemented *polynomial approximation* and *thin-plate spline interpolation* to do the correction.

2.2.1 Radiometric Correction

Radiometric distortion is modeled by a series of concentric ellipses. The conventional polynomial approximation is modified to be able to inverse this model (See details in Appendix 7.4). Fig. 5 shows two real images (a grid template and a flat frame) captured from our imaging system, and the corrected results. It is apparent that the concentric ellipse effect disappeared in the corrected images.

Table 1 is a rough measurement on how the correction algorithm works. Since the correction algorithm is to correct the degraded brightness with central pixel brightness as the reference, the mean brightness of the corrected image should be always larger than that of the measured image. On the other hand, for the grid template, the corrected image show better contrast between white pixel areas and dark grid lines, therefore, the variance of the corrected image is larger than that of the measured image; while for the flat frame image, the variance is much less than that of the measured image since the corrected image has a homogeneous brightness distribution.

TABLE 1. Mean and Variance comparison.

		mean (μ)	variance (σ^2)
grid template	measured image	147.24	6474.32
	corrected image	183.57	9601.04
flat frame	measured image	188.02	285.42
	corrected image	251.60	52.67

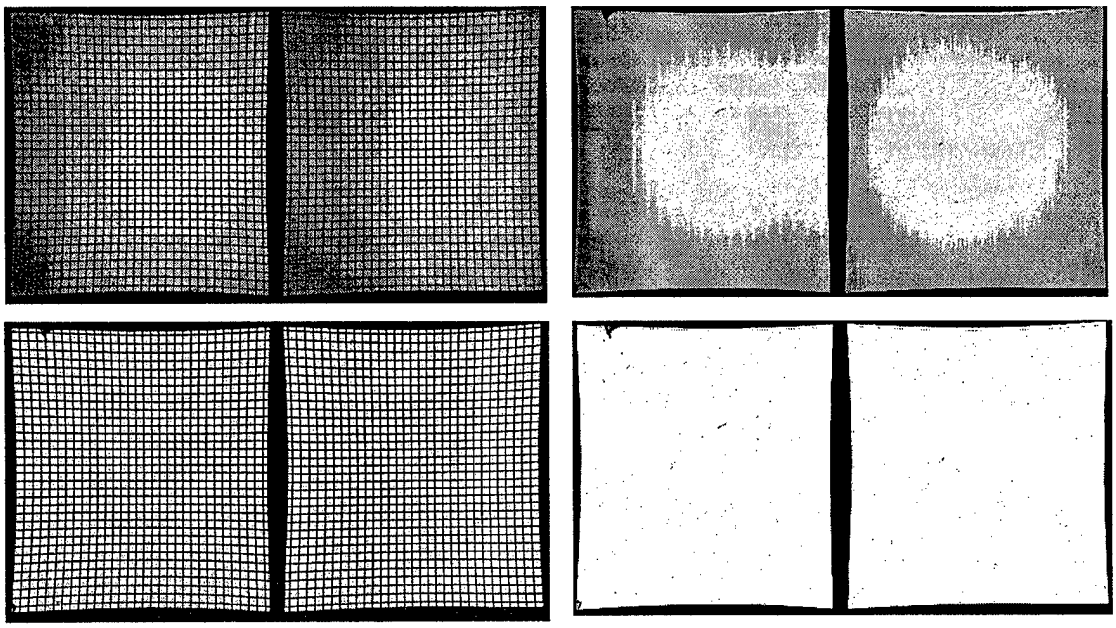


Figure 5. Results from radiometric correction of a grid template (left) and a flat frame (right).

We also use histograms and profiles to evaluate the correction results. Fig. 6 shows histograms of measured images and corrected ones. We can see that the histogram of the measured grid template has two peaks: one is from dark grid lines; the other smoother peak is from white pixel areas which shows the shape of a Gaussian because of radiometric distortion. The histogram of the corrected grid template also has two peaks: one is from dark grid lines; the other peak is an impulse (Dirac delta function) instead of a Gaussian, which indicates the successful correction. Histograms of the measured flat frame and the corrected frame show similar characteristics as those of the grid template except that the smaller peak is from the boundary of the image area.

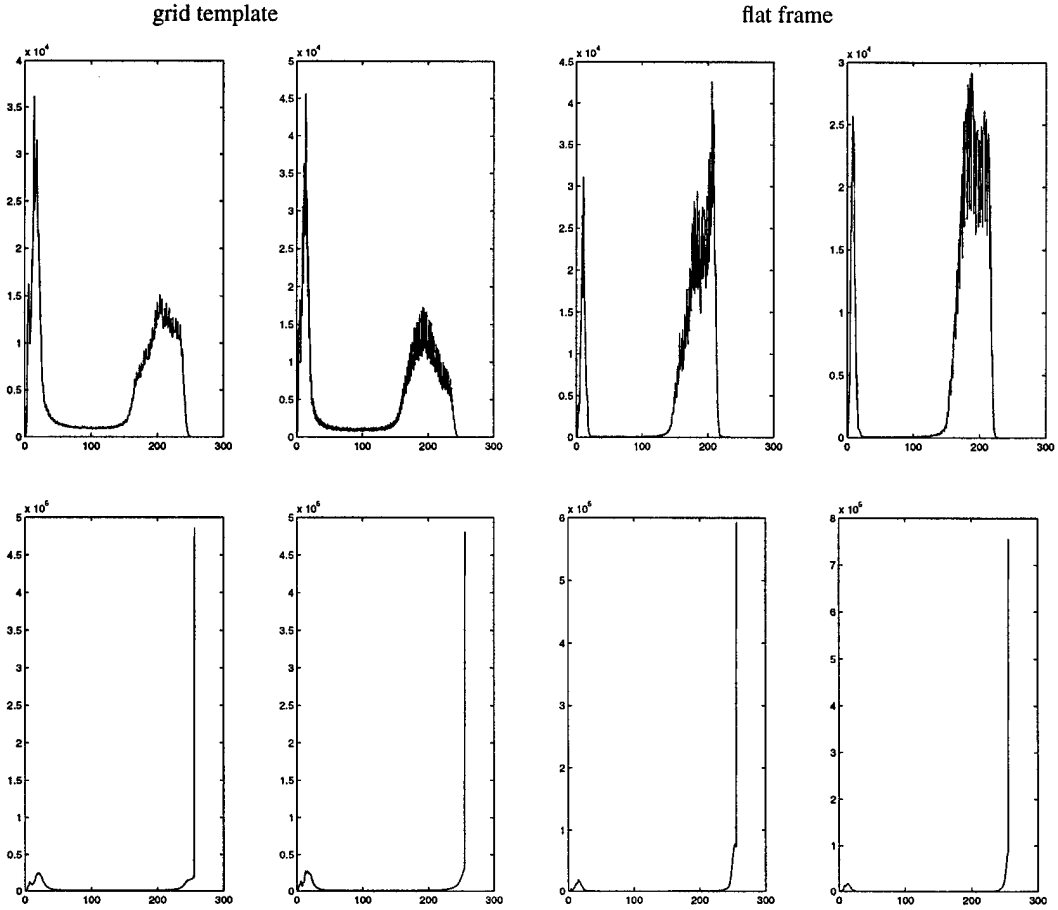


Figure 6. Comparison of histograms of the measured images (top), and the corrected images (bottom). In each figure, the left plot is for the image from the left sensor, and the right plot is that from the right sensor.

Fig. 7 is a set of profiles used to measure performance of the algorithm in further detail. These profiles are from the same columns/rows of the measured grid template image and the corrected image. Profiles on the left column are from the measured image which show a Gaussian envelope; while the profiles to the right are from the corrected image that show a line envelope.

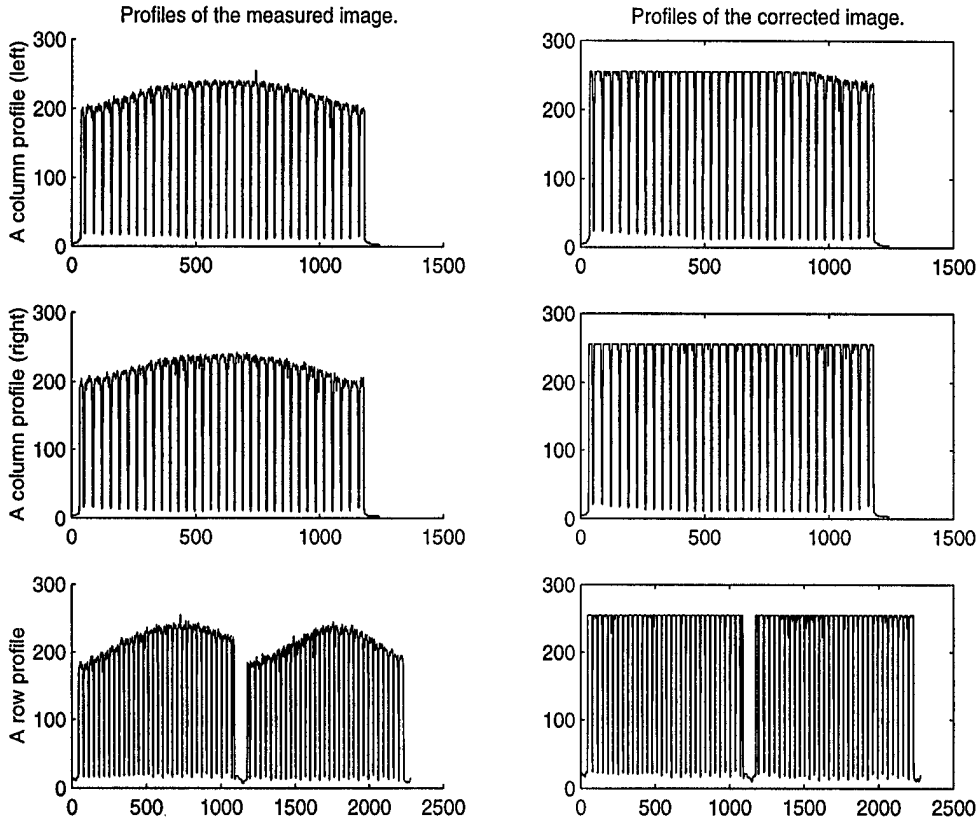


Figure 7. Comparison of profiles from the same columns and row of the measured grid template image and the corrected image.

2.2.2 Geometric Correction

Both polynomial approximation and TPS interpolation are implemented to correct pincushion distortion. The difference between *approximation* and *interpolation* lies in the fact that approximation methods generate transformations that map all the control points *close* to their correspondence, so that the summation of displacements achieves a global minimum; whereas interpolation methods produce transformations where all the control points can be mapped to their correspondence *exactly* [Daview and Samuels 96] (Appendix 7.5).

In the following experiments, control points are chosen to be equally distributed across the whole image. A grid template is generated as a gold standard for comparison. Fig. 8. displays the corrected results by both methods with different numbers of control points chosen. It is not that apparent as to which graph shows better result. Therefore, we need to use some quantitative measurement to compare the results.

Two numbers are calculated to compare the quality of the corrected images: the error rate (ε) and the cross correlation coefficient (ρ).

The error rate ε is defined by Eq. (1). Since the template is a binary image with the brightness of grid lines as 0s, and other places 1s, the definition of the error rate in Eq. (1) is the same as the *mean square error* (MSE).

$$\varepsilon = \frac{\text{sum of misaligned pixels}}{\text{total number of pixels}} \tag{1}$$

The cross correlation coefficient ρ is defined as Eq. (2).

$$\rho_{x,y} = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} = \frac{E[(x-E[x])(y-E[y])]}{\sqrt{E[x-E[x]]E[y-E[y]]}} \quad (2)$$

Fig. 9 and Fig. 10 are two plots of error rate comparison and cross correlation coefficients comparison.

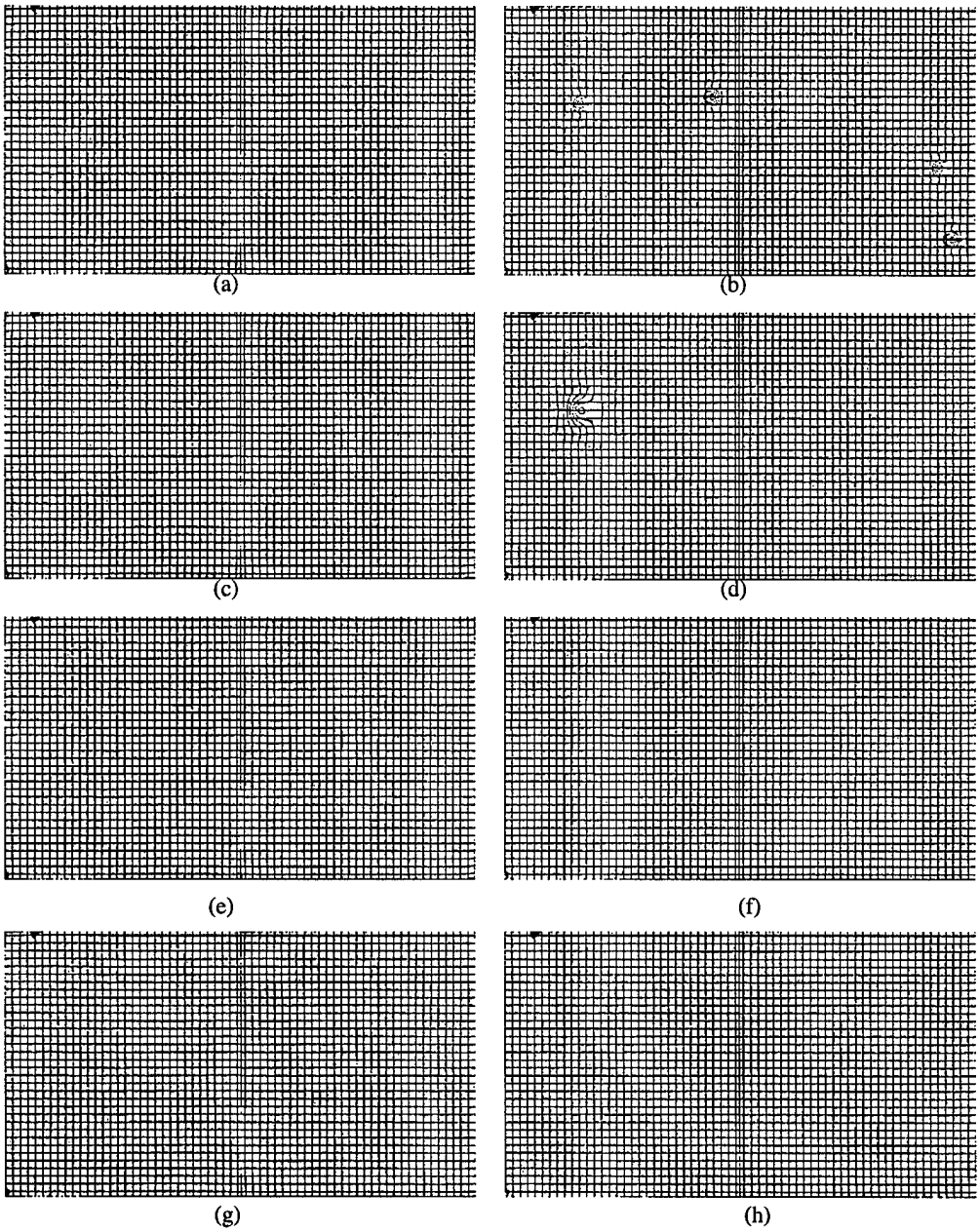


Figure 8. Correction results from both polynomial approximation (left column) and TPS interpolation (right column) with respect to different numbers of control points. From top to bottom: 33 x 33, 17 x 17, 9 x 9, and 5 x 5. The artifacts in (b) and (d) are explained on page 9.

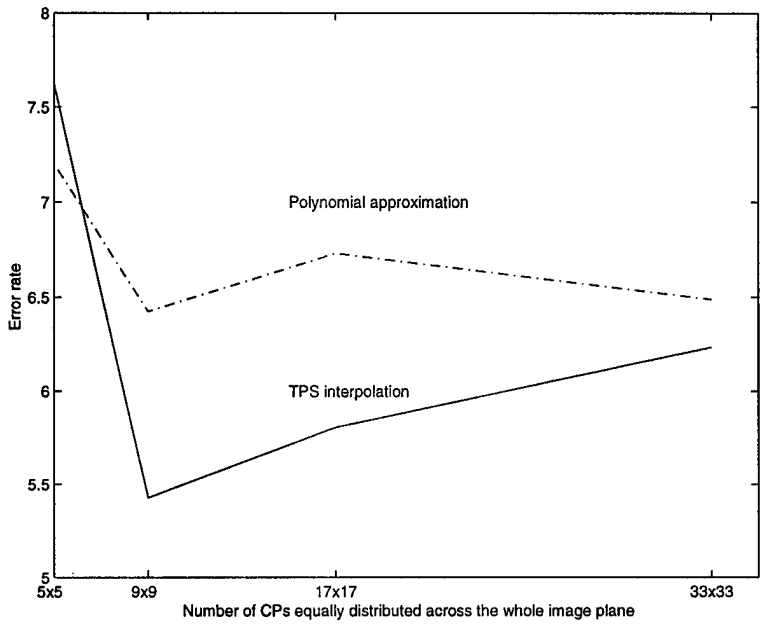


Figure 9. Comparison of error rates of polynomial approximation and TPS interpolation as a function of the number of control points.

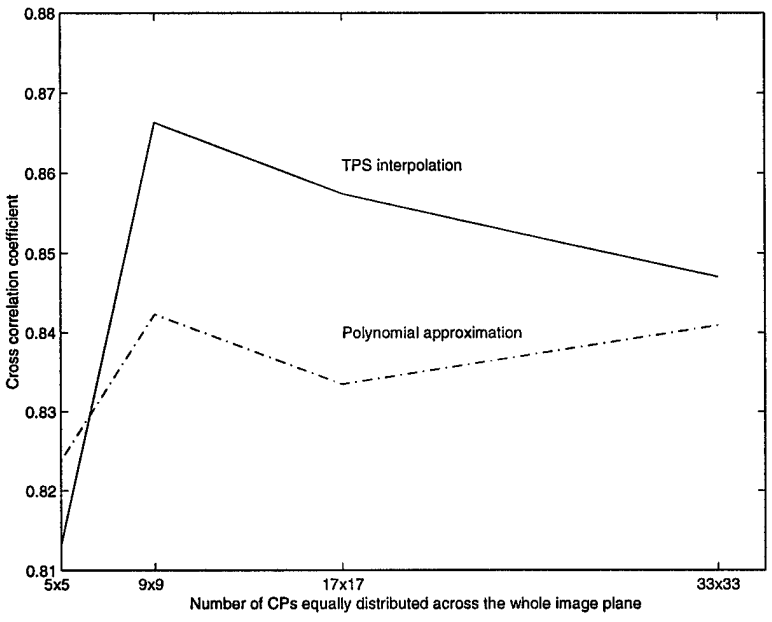


Figure 10. Comparison of cross correlation coefficient between different correction results and the template.

From the above comparisons, we draw the following conclusions: (1) In general, TPS interpolation performs better than polynomial approximation. Except that with 5 x 5 control points, TPS has a lower error rate and a higher cross correlation coefficient than those of the polynomial approximation. (2) The quality of the corrected image relates to the number of control points being chosen. Surprisingly, it is not true that the more control points, the less the error rate. The corrected image has the lowest error rate and highest correlation coefficient with 9 x 9 control points. The reason is stated as follows:

- The interpolation approach can achieve *exact* mapping based on the assumption that the control point positions are known *exactly.* In real applications, however, the positions of the control points in the measured image can only be determined *approximately.* Therefore, the more control points chosen, the more the potential error might accumulate, resulting in the severe local distortion in the corrected images when choosing 33 x 33 control points, and 17 x 17 control points, as can be observed in Fig. 8. More local distortions exist in the correct image with 33 x 33 control points than that with 17 x 17 control points.
- For approximation approach, the more control points chosen, the more tendency the mapping functions show towards discontinuity. In other words, the mapping function might fit well for the control points, but not globally for all the points in the image. Therefore, unless we choose enough control points (e.g. all the points in the image), or we will get more errors when continue increasing the number of control points.

2.3 Missing Data Estimation

Based on the degradation analysis in Appendix 7.6, the image restoration problem can be formulated as Fig. 11. An original image f is slightly blurred by a point spread function (PSF) h from the X-ray source. The blurred image is further corrupted by fixed noise (n) and defects (d) from the CCD detector. g is the so called *measured image.*

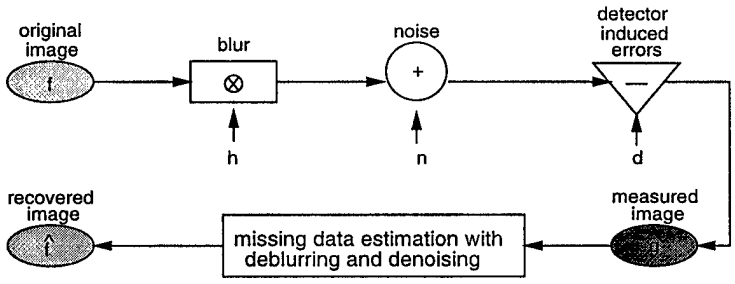


Figure 11. System model.

The image formation process can be expressed by Eq. (3), where $\otimes$ stands for the convolution operator.

$$g = f \otimes h + n - d \tag{3}$$

where we assume that the PSF is already known, the noise distributes as independent Gaussian, and the positions of missing data can be identified.

We propose two approaches to estimate missing data along with deblurring and denoising: the *consistency method* using separable deblurring, and a maximum *a-posteriori* probability (MAP) method, in which the optimization is solved using Mean Field Annealing (MFA).

The basic idea behind both of our approaches is to make use of the point spread function (PSF): before a pixel is missed, it has already distributed its information to its neighbors through the effect of blur. Fig. 12 illustrates this process, where each block indicates a pixel, and the blocks marked with "0"s are the missing pixels. The figure shows that before pixel C is missed, it spreads its information to the neighbors $g1$, $g3$, $g4$, $g6$, $g7$, and $g9$.

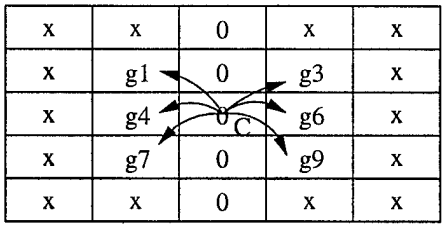


Figure 12. Information distribution by blur.

If we can obtain a measurement of the PSF, in theory, we can reconstruct the missing data. Regularization theory (Appendix 7.7) is used to convert an ill-posed problem to a well-posed one. The consistency method using separable deblurring (Appendix 7.9) constructs a well-posed problem by putting several restrictions on the image data and the blur kernel. The MAP method (Appendix 7.10) is actually a Bayesian interpretation of regularization problems, which is solved by a global optimization technique, called mean field annealing (MFA).

Experiments are carried on both synthetic images and real images. System distortions and degradations are characterized in Appendix 7.8.

The synthetic image are generated by simulating the image formation process: (1) blur the original image (f) by convolving f with a finite blur kernel h calculated in Appendix 7.8; (2) insert Gaussian white noise with zero mean and standard deviation equal to 3.9, which is obtained in Appendix 7.8; and (3) remove one column of data to generate the measured image g. Five synthetic images are generated showing different degrees of smoothness and different image properties. These images are a piecewise constant image, a piecewise linear image, a piecewise quadratic image, a sinusoid image, and a mammogram (Fig. 13).

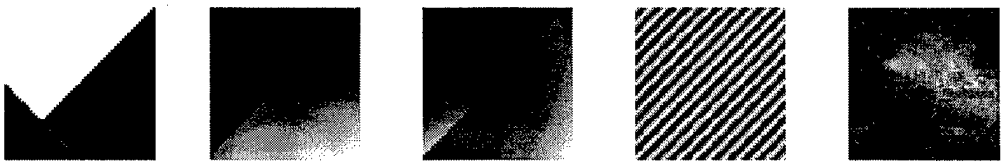


Figure 13. Five synthetic images with different properties. From left to right: piecewise constant, piecewise linear, piecewise quadratic, sinusoid, and mammogram.

We conduct experiments by relaxing the following two assumptions one by one in order to investigate the effect of each assumption to the quality of the restored image:

1) the blur kernel is exactly known; and

2) noise is not inserted.

Experimental results from both the consistency method (using integer criterion and neighbor least square error criterion) and the MFA are exhibited and compared based on line profiles and column profiles.

2.3.1 Results Based on Complete Set of Assumptions

With the complete set of assumptions (the blur kernel is exactly known, the blur kernel is separable, the original image is of integer type, and no noise is inserted) satisfied, the consistency method using integer criterion can recover the missing column exactly from the measured image. The consistency method using neighbor least square error (NLSE) criterion shows better results than the MFA method in deblurring, as

shown in Fig. 14. To compare the results in detail, we use column/row profile comparisons, which are exhibited in Fig. 15.

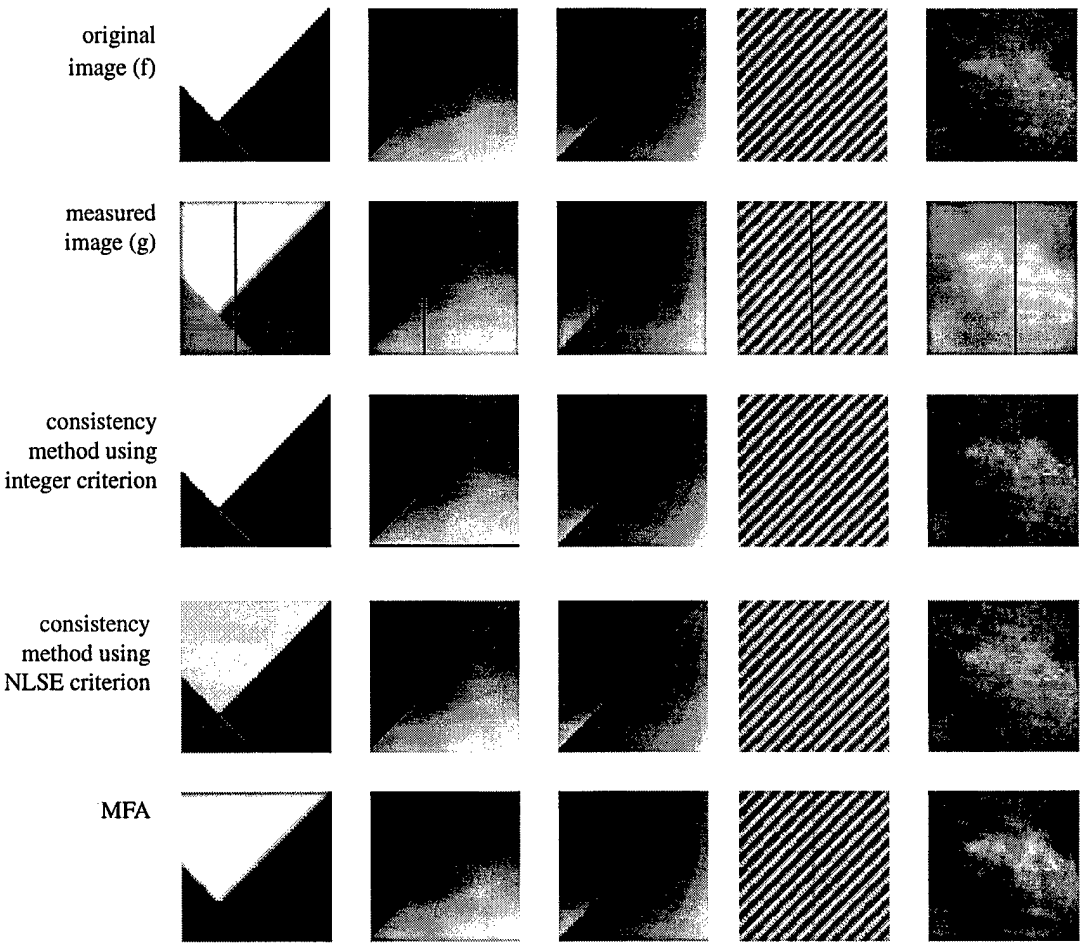


Figure 14. Restoration of integer images blurred by an exactly known kernel.

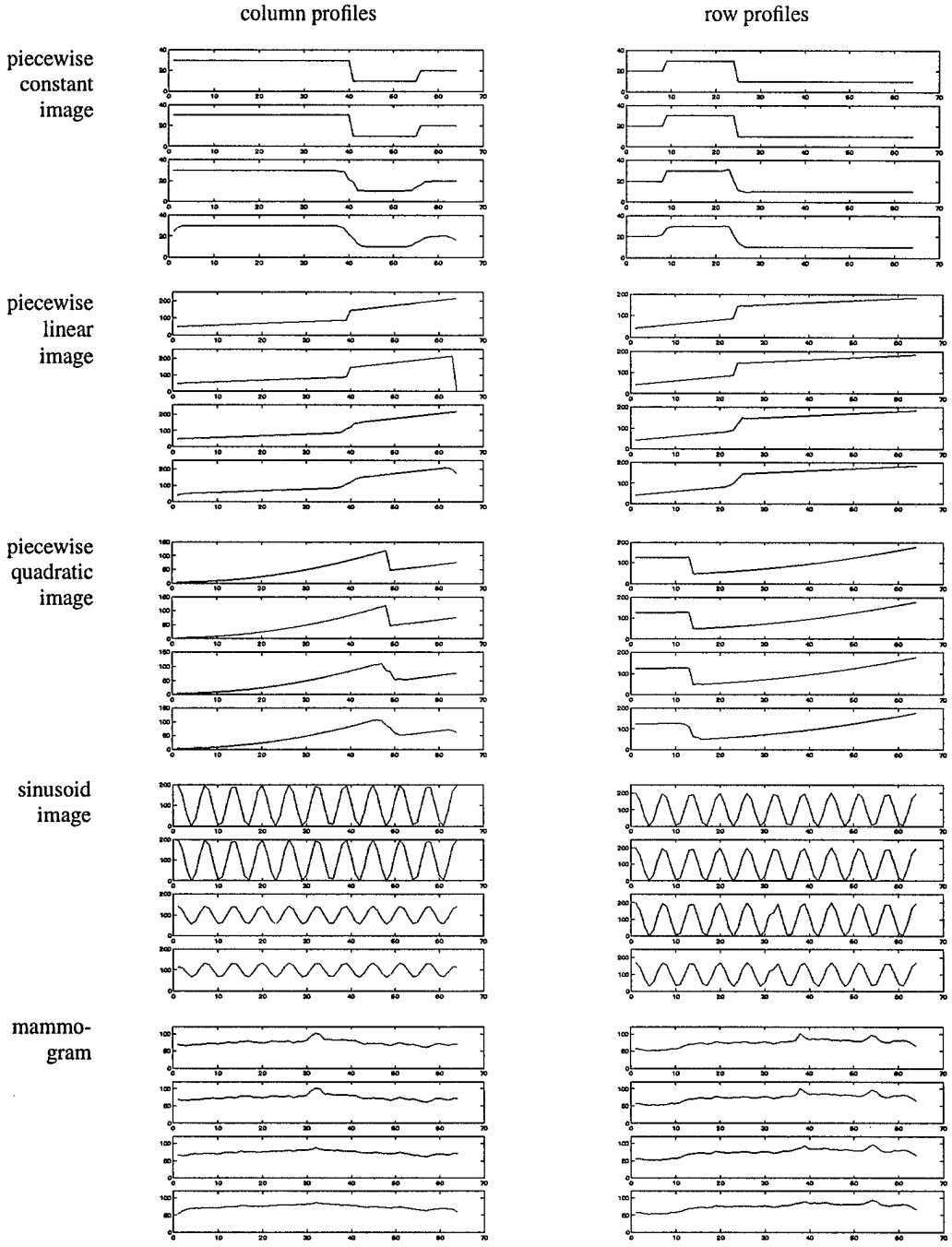


Figure 15. Profile comparisons of restoration results from the five synthetic images. In each plot, from top to bottom: profile from the original image, restored images by consistency (int), consistency (NLSE), and MFA.

A cross-correlation comparison between the original image and the restored one using different methods is another way to compare the algorithm performance, as shown in Fig. 16.

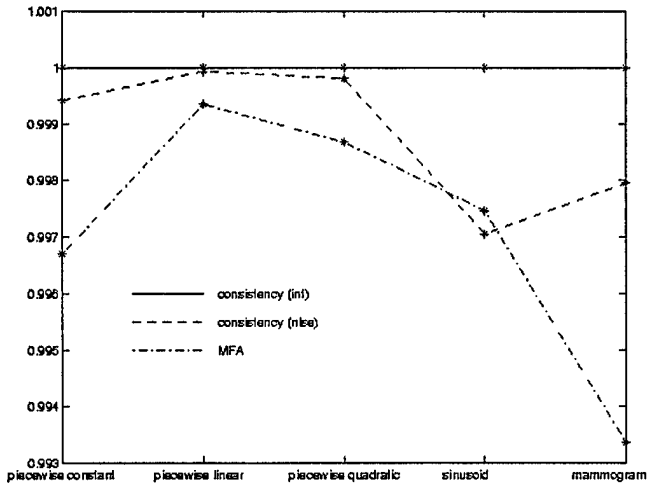


Figure 16. Cross-correlation comparison.

The consistency method with integer criterion always has the highest cross-correlation coefficient, which is 1. The consistency method with NLSE criterion shows better correlation than the MFA method most of the time except for sinusoid images.

2.3.2 Relaxing the Assumption of Perfect Blur Kernel Estimation

We use blur kernel h to blur the original image, but use $h + \varepsilon$ to restore the measured image, where ε represents the error added to a blur kernel. The accuracy of an estimated blur kernel is measured by correct number of digits. For example, $\varepsilon = 1\%$ means the 2nd digit of elements in the blur kernel is not correct, and $\varepsilon = 0.1\%$ means the error shows up in the 3rd digit. Therefore, if h = [0.0625 0.1250 0.625 0.1250 0.0625], a possible estimation when $\varepsilon = 1\%$ is h = [0.07 0.13 0.63 0.13 0.07], where the 2nd digit is not accurate. Fig. 17 shows restoration results of consistency method using NLSE criterion and MFA method at different error rate ε (0.1% and 1%), Fig. 18 is a column/row profile comparison, and Fig. 19 is the cross-correlation comparison.

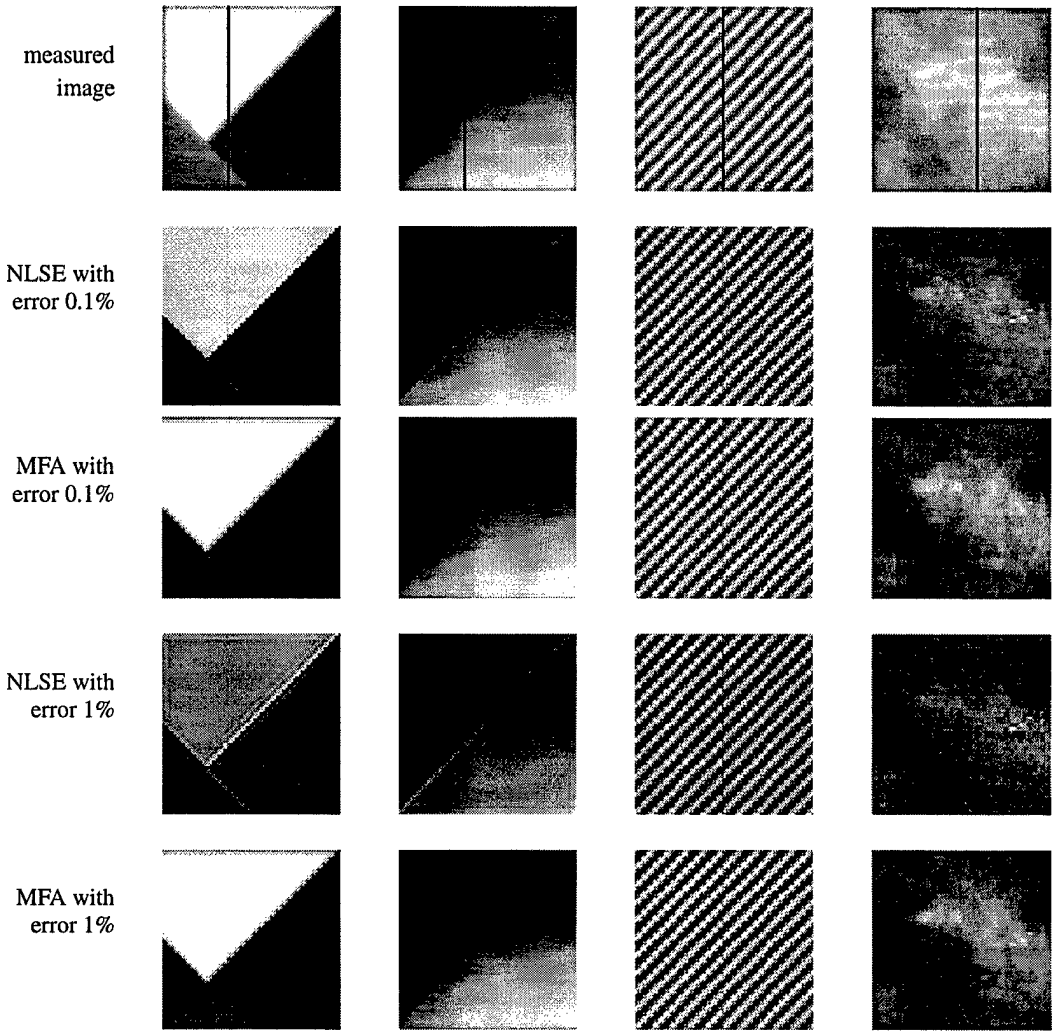


Figure 17. Restoration results with an inaccurate blur kernel.

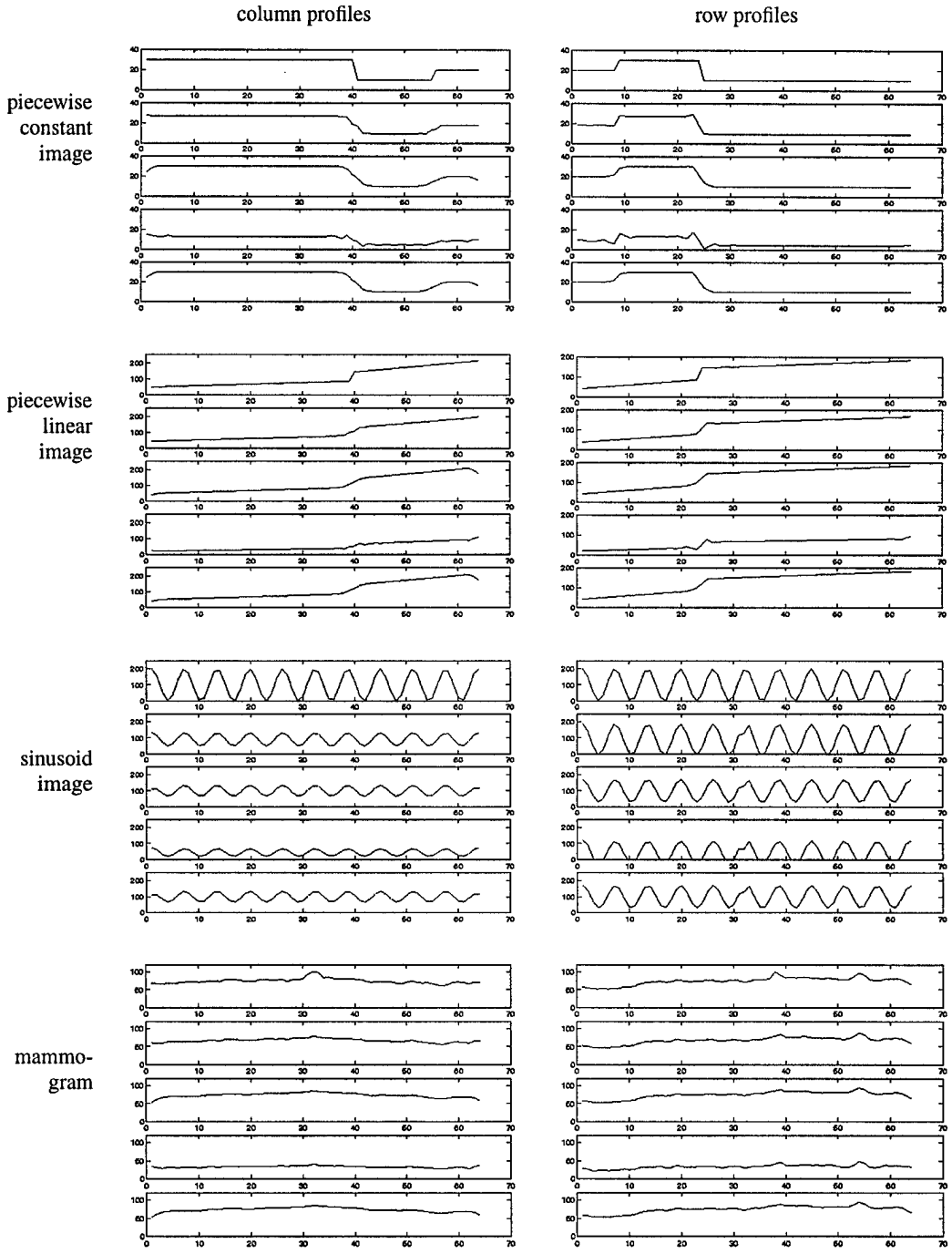


Figure 18. Profile comparisons of restoration results. In each plot, from top to bottom: profile from the original image, restored images by consistency (NLSE) with error 0.1%, by MFA with error 0.1%, by consistency (NLSE) with error 1%, and by MFA with error 1%.

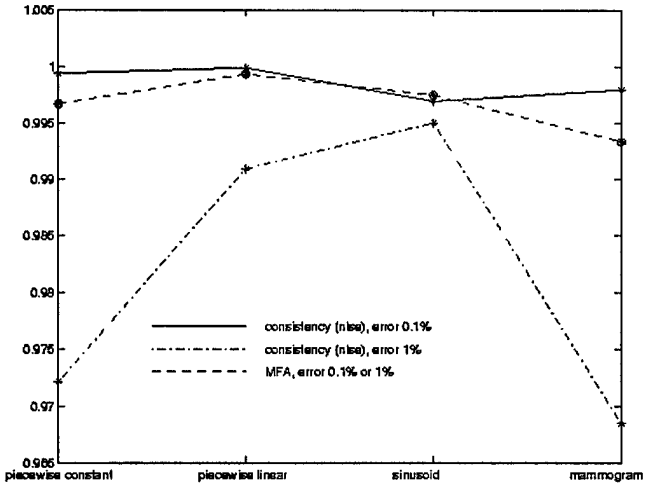


Figure 19. Cross-correlation comparison.

From Fig. 19, we can see that the MFA method is not sensitive to very small errors in the blur kernel since the correlation coefficients do not change when increasing the error rate from 0.1% to 1%. However, the coefficients of the consistency method with NLSE criterion drops when increasing the error. When error is on the 2nd digit ($\varepsilon = 0.1\%$), the consistency method provides better correlation than the MFA method except for the sinusoid image.

To summarize, when the estimated blur kernel is in high accuracy (two or more than two digits are correct), the consistency method performs better than the MFA method. However, the consistency method is more sensitive than MFA when error in the estimated kernel is increased.

2.3.3 Relaxing the Noise-Free Assumption

Gaussian white noises $n_1 \sim N(0, 0.01^2)$, $n_2 \sim N(0, 0.1^2)$ and $n_3 \sim N(0, 1)$ are inserted to the blurred image, and one column of data are deleted. Restoration results on the piecewise constant image are shown in Fig. 20. Column/row profiles are compared in Fig. 21 and Fig. 22.

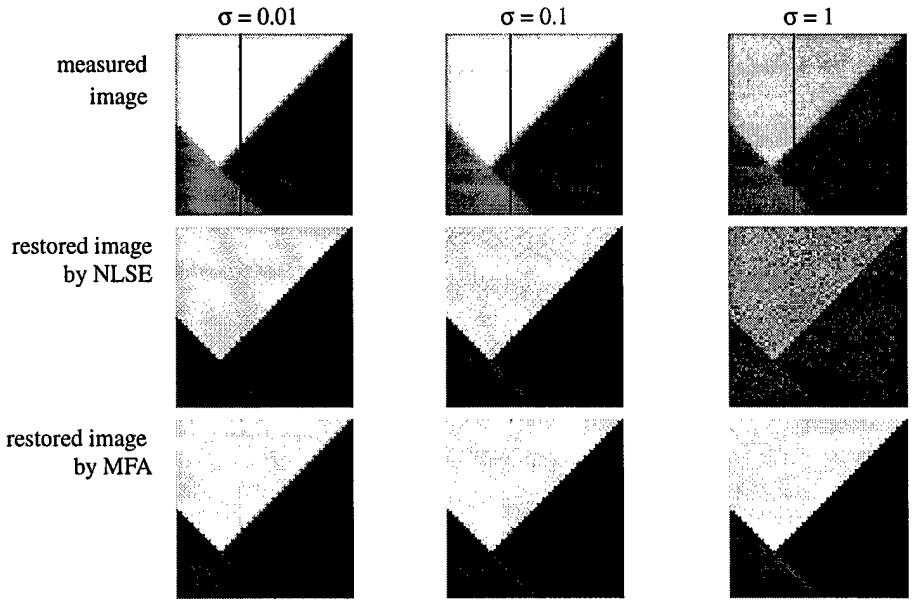


Figure 20. Restoration results with noise inserted in the blurred image.

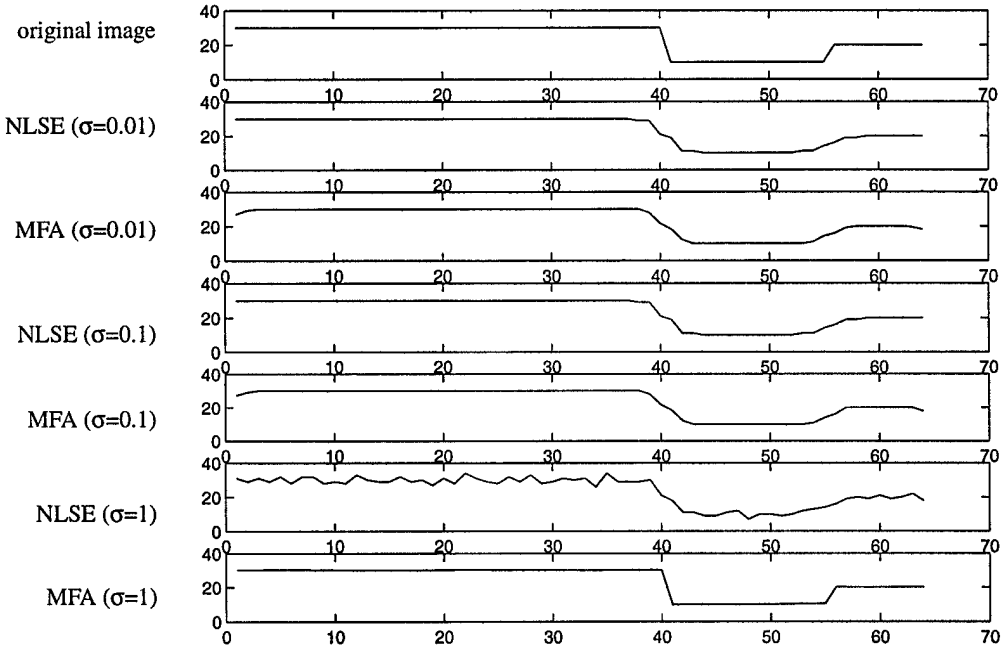


Figure 21. Column profiles comparison.

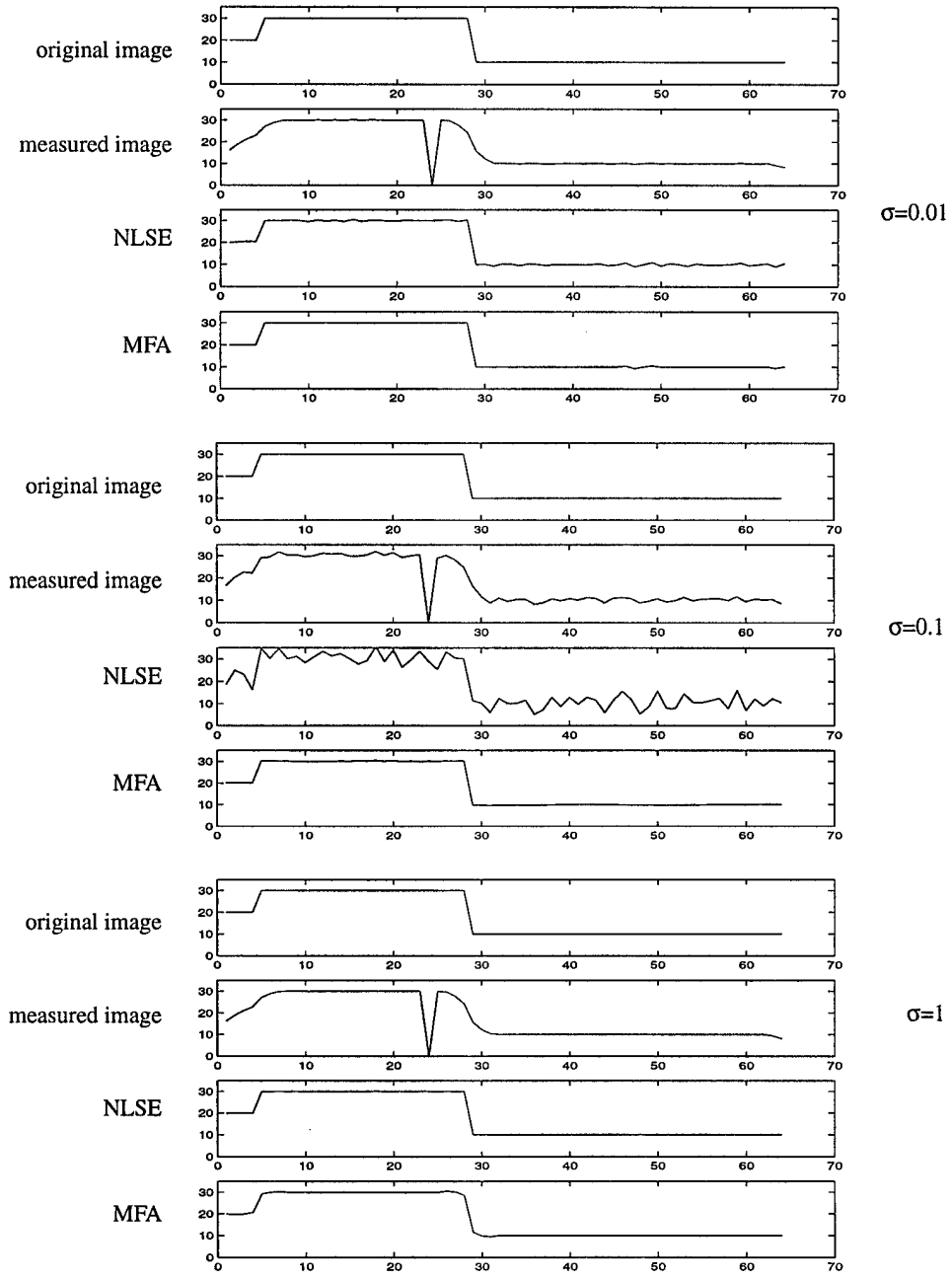


Figure 22. Row profiles comparison.

The results show us that when noise is very small, the consistency method performs better than the MFA method, it achieves a higher correlation coefficient (0.998269) than the MFA method (0.998084). However, when increasing the standard deviation of the inserted noise, the consistency method is much more disturbed than the MFA method.

2.3.4 Restoration Results of Real Image

Fig. 23 exhibits a stripe of the grid template image (Fig. 44), and its restoration results using both the consistency method and the MFA method.

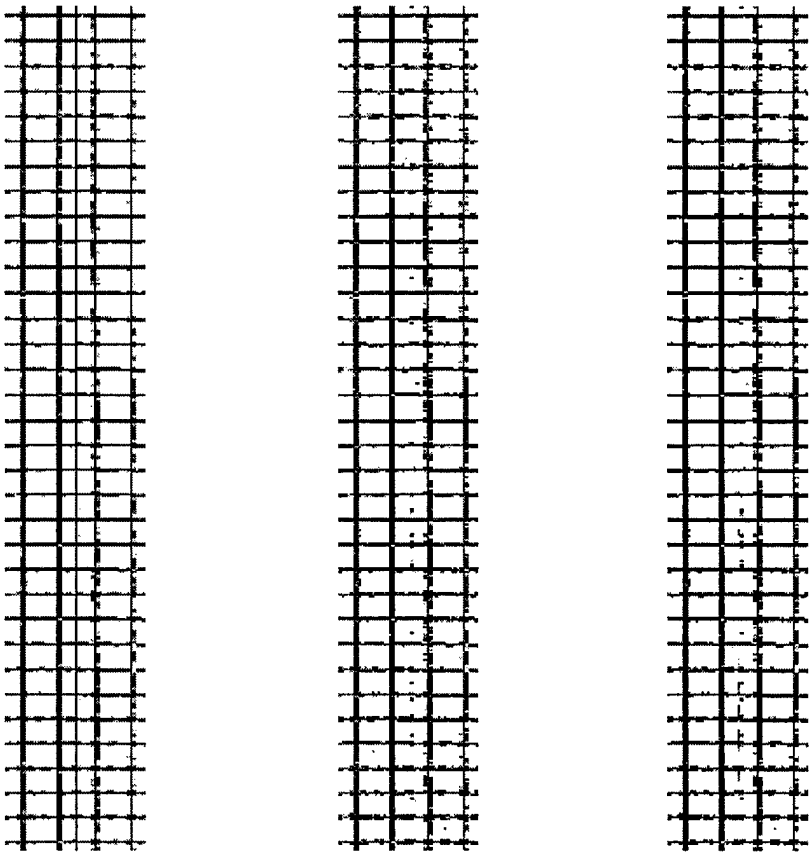


Figure 23. Restoration results of a stripe of grid template image with flash light source. From left to right: the measured image, restored image by NLSE, restored image by MFA.

The restored image from the consistency method is very similar to that from the MFA method, but the consistency method is more efficient according to the complexity analysis in Appendix 7.11.

To summarize, the consistency method works better when noise is very small, and blur kernel is estimated with high accuracy. Performance of the consistency method may be largely affected by perturbations in the estimated blur matrix, or by noise, with perturbations in noise playing a more important role. The MFA method behaves more stable for large noise or inaccuracy in the blur kernel estimation.

2.4 Original Contribution

The original contribution of this research work includes the use of conditioning theory in the analysis of image restoration problem, the development of consistency method with separable deblurring, and the use of thin-plate spline interpolation algorithm in geometric correction.

2.4.1 Conditioning Analysis of Missing Data Estimation

While there have been many articles in the literature discussing various image restoration algorithms, few [Forbes and Anh 94][Milinazzo et al. 87] pay attention to quantify the ill-conditioning of the blur kernel. The most important contribution of this research work is that we use conditioning theory to analyze image

restoration problems, which has helped us obtain an in-depth view of how algorithms might behave theoretically and how perturbation of each parameter might affect the solution. We also use the conditioning analysis to guide our algorithm design and parameter selection. The conclusions we draw from the conditioning analysis are well demonstrated by the experimental results.

2.4.2 Consistency Method with Separable Deblurring

Neither separable convolution nor the idea of converting a convolution operation into a matrix multiplication is new. These techniques have been addressed very early, back to 70s [Andrews and Hunt 77]. Our contribution is that we first take advantage of the separability property of some blur kernel (like Gaussian), and convert the separable convolution operation into a separable matrix multiplication. By inserting several restrictions on the data, we can construct the separated circulant matrix and be able to demonstrate that the condition number of such matrices is very small, so that the problem is well-conditioned. In this way, we convert the ill-posed image restoration problem to a well-posed one.

By using a backward-stable algorithm, we are able to restore the missing data with 100% accuracy assuming the defects only come from blur and the original image is of integer type.

The beauty of this algorithm lies in its simplicity and computational efficiency.

2.4.3 TPS in Geometric Correction

TPS interpolation is a conventional method for interpolation. We first apply it to correct the geometric distortion. No other published literature has used this method under this scenario. By using TPS interpolation, we gain better performance in the correction results than the traditional polynomial method.

3. List of Publications

- Snyder, W.E., Qi, H., Elliott, R.L., Head, J.F. and Wang, C.X. (1999). "Increasing the effective resolution of thermal infrared images". Accepted by *IEEE Engineering in Medicine and Biology Magazine.*
- Qi, H. and Snyder, W.E. (1999). "Content-based image retrieval in picture archiving and communications systems". *Journal of Digital Imaging*, 12(2): 81-83, Suppl 1, May.
- Qi, H. and Snyder, W.E. (1999). "Lesion detection and characterization in digital mammography by Bezier histograms". *SPIE Medical Imaging*, San Diego, CA, February.
- Snyder, W.E., Qi, H. and Sander, W. (1999). "A hexagonal coordinate system". *SPIE Medical Imaging*, San Diego, CA, February.
- Qi, H., Snyder, W.E. and Bilbro, G.L. (1998). "Missing data estimation by separable deblurring". *Proceedings for the IEEE International Joint Symposia on Intelligence and Systems*, pp. 348-353, Rockville, MD, May.
- Qi, H., Snyder, W.E. and Bilbro, G.L. (1998). "Comparison of mean field annealing and multiresolution analysis in missing data estimation". *Computer Vision - ACCV'98: Third Asian Conference on Computer Vision*, v1, pp722-729, Hong Kong, China, January 8-10.
- Qi, H., Snyder, W.E. and Bilbro, G.L. (1997). "Using mean field annealing to solve anisotropic diffusion problems". *IEEE International Conference on Image Processing*, v3, pp. 352-355, Santa Barbara, CA.

4. Scientific Personnel

- Dr. Wesley E. Snyder, Professor of the Department of Electrical and Computer Engineering, NCSU
- Dr. Griff L. Bilbro, Associate Professor of the Department of Electrical and Computer Engineering, NCSU
- Dr. Hairong Qi, Degree earned on August 12, 1999 at the Department of Electrical and Computer Engineering, NCSU

5. Reports of Inventions

N/A

6. Bibliography

[Abdelqader et al. 92] Abdelqader, I., Rajala, S., Snyder, W.E., and Bilbro, G. (1992). "Energy minimization approach to motion estimation using mean field annealing". *Signal Processing*, July.

[Aghdasi et al. 94] Aghdasi, F., Ward, R.K., and Palcic, B., "Restoration of mammographic images in the presence of signal-dependent noise". *State of the Art in Digital Mammographic Image Analysis*, pp. 42-63, World Scientific, 1994.

[Andrews and Hunt 77] Andrews, H.C. and Hunt, B.R. (1977). *Digital Image Restoration*, Prentice Hall.

[Beier et al. 92] Beier, J., Oswald, H. and Fleck, E. (1992). "Edge detection for coronary angiograms: error correction and impact of derivatives", *Proceedings of the 18th Annual Conference on Computers in Cardiology*, pp. 513-516.

[Bender et al. 96] Bender, W., Gruhl, D., Morimoto, N. and Lu, A. (1996). "Techniques for data hiding", *IBM Systems Journal*, 35(NOS 3&4): 313-335.

[Besag 74] Besag, J. (1974). "Spatial interaction and statistical analysis of lattice systems". *J. Roy. Stat. Soc. Lond. B*, v36: 192-225.

[Besag 86] Besag, J. (1986). "On the statistical analysis of dirty pictures". *J. R. Statist. Soc. B*, 48 (3): 259-302.

[Beynon and Lamb 80] Beynon, J.D.E. and Lamb, D.R. (1980). *Charge-coupled devices and their applications*, McGraw-Hill Book Company (UK) Limited, Maidenhead, Berkshire, England.

[Bilbro and Snyder 88a] Bilbro, G.L. and Snyder, W.E. (1988). "Image restoration by mean field annealing". *Advances in Neural Network Information Processing Systems*.

[Bilbro and Snyder 88b] Bilbro, G.L. and Snyder, W.E. (1988). "Fusion of range and luminance data". *IEEE Symposium on Intelligent Control*, Arlington, August.

[Bilbro et al. 89] Bilbro, G.L., Mann, R., Miller, T., Snyder, W.E., et al. (1989). "Optimization by mean field annealing". *Advances in Neural Information Processing Systems*. Morgan-Kauffman, San Mateo.

[Bilbro and Snyder 89] Bilbro, G.L. and Snyder, W.E. (1989). "Range image restoration using mean field annealing". *Advances in Neural Information Processing Systems*. Morgan-Kauffman, San Mateo.

[Bilbro and Snyder 90] Bilbro, G.L. and Snyder, W.E. (1990). "Applying mean field annealing to image noise removal". *J. of Neural Network Computing*, pp. 5-17, Fall.

[Bilbro and Snyder 91] Bilbro, G.L. and Snyder, W.E. (1991). "Optimization of functions with many minima". *IEEE Transactions on Systems, Man, and Cybernetics*, 21(4): 840-849, July/August.

[Bilbro et al. 92] Bilbro, G.L., Snyder, W.E., Garnier, S.J. and Gault, J.W. (1992). "Mean field annealing: a formalism for constructing GNC-like algorithms". *IEEE Trans. on Neural Networks*, 3 (1): 131-138, January.

[Bilbro et al. 98] Bilbro, G.L., Hall, L.C., Clements, M. and etc. (1998). "Convolution, deconvolution, and mean field annealing for analog VLSI". *IEEE Trans. on Circuits and Systems*, Part II, 46(2): 120-128, February.

[Blake and Zisserman 87] Blake, A. and Zisserman, A. (1987). *Visual Reconstruction*. The MIT Press, Cambridge.

[Bookstein 78] Bookstein, F.L. (1978). *The Measurement of Biological Shape and Shape Change*, Springer-Verlag. Editor: Levin, S., Lecture Notes in Biomathematics, Berlin Heidelberg.

[Bookstein 89] Bookstein, F.L. (1989). "Principal warps: thin-plate splines and the decomposition of deformations", *IEEE PAMI*, 11(6): 567-585, June.

[Butler and Pierson 91] Butler, D.A. and Pierson, P.K. (1991). "A distortion-correction scheme for industrial machine-vision application", *IEEE Transactions on Robotics and Automation*, 7(4): 546-551, August.

[Chang et al. 95] Chang, S.G., Cvetkovic, Z. and Vetterli, M. (1995). "Resolution enhancement of images using wavelet transform extrema extrapolation". *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (IEEE)*, Part 4, pp. 2379-2382, Detroit, MI.

[Chen and de Figueiredo 93] Chen, G. and de Figueiredo, R.J.P. (1993). "A unified approach to optimal image interpolation problems based on linear partial differential equation models". *IEEE Transactions on Image Processing*, 2(1): 41-49, January.

[Cox et al. 97] Cox, I.J., Kilian, J., Leighton, T.F. and Shamoon, T. (1997). "Secure spread spectrum watermarking for multimedia", *IEEE Transactions on Image Processing*, 6(12): 1673-1687, December.

[Daview and Samuels 96] Daview A. and Samuels P. (1996). *An Introduction to Computational Geometry for Curves and Surfaces*, Oxford University Press, Oxford Applied Mathematics and Computing Science Series, New York.

[Delbrueck 93] Delbrueck, T. (1993). "Silicon retina with correlation-based, velocity-tuned pixels". *IEEE Transactions on Neural Networks*, 4(3): 529-541.

[Demoment 89] Demoment, G. (1989). "Image reconstruction and restoration: overview of common estimation structures and problems". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 37 (12): 2024-2036, December.

[Frei 83] Frei, W. (1983). *Digital Image Processing*. Course 411, Course Materials from the Learning Tree, Integrated Computer Systems.

[Forbes and Anh 94] Forbes, K. and Anh, V.V. (1994). "Condition of system matrices in image restoration". *J. Opt. Soc. Am. A*, 11 (6): 1727-1735, June.

[Geman and Geman 84] Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". *IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI*-6 (6): 721-741, November.

[Gerchberg 74] Gerchberg, R.W. (1974). "Superresolution through error energy reduction". *Optical Acta.*, 21(9): 709-720.

[Hammersley and Handscomb 64] Hammersley, J.M. and Handscomb, D.C. (1964). *Monte Carlo Methods*. New York: Wiley.

[Hiriyannaiah et al. 89] Hiriyannaiah, H.P., Bilbro, G.L. and Snyder, W.E. (1989). "Restoration of piecewise-constant images by mean-field annealing". *J. Opt. Soc. Am. A*, 6 (12): 1901-1912, December.

[Hussain 97] Hussain, A. (1997). *A-Posteriori Estimation of the Point Spread Function to Sub-Pixel Accuracy*, MS thesis, North Carolina State University, Raleigh, NC.

[Katsaggelos 91] Katsaggelos, A.K. (1991). *Digital Image Restoration*, Springer-Verlag.

[Lemeire 75] Lemeire, F. (1975). "Bounds for condition numbers of triangular and trapezoid matrices". *BIT*, 15: 58-64.

[Li 95] Li, S.Z. (1995). *Markov Random Field Modeling in Computer Vision*, Springer.

[Marsi and Carrato 95] Marsi, S. and Carrato, S. (1995). "Neural network-based image segmentation for image interpolation". *Proceedings of the 5th IEEE Workshop on Neural Networks for Signal Processing (NNSP'95)*, pp. 388-397, Cambridge, MA.

[Meinguet 84] Meinguet, J. (1984). "Surface spline interpolation: basic theory and computational aspects", *Approximation Theory and Spline Functions*, pp. 127-142, Kluwer Academic Publishers.

[Mendis and Pain 93] Mendis, S.K., Pain, B. (1993). "Low-light-level image sensor with on-chip signal processing". SPIE vol. 1952, pp. 23-33.

[Milinazzo et al. 87] Milinazzo, F., Zala, C. and Barrodale, I. (1987). "On the rate of growth of condition numbers for convolution matrices". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-35 (4): 471-475, April.

[Nashed 76] Nashed, M.Z. (1976). "Aspects of generalized inverses in analysis and regularization". In *Generalized Inverses and Applications*, ed. by M.Z. Nashed, Academic, New York.

[Nicholas 87] Nicholas, H.J. (1987). "A survey of condition number estimation for triangular matrices". *SIAM Review*, 29 (4): 575-596, December.

[Papoulis 75] Papoulis, A. (1975). "A new algorithm in spectral analysis and band-limited extrapolation". *IEEE Transactions on Circuit and System*, 22: 735-742, September.

[Parker et al. 83] Parker, J.A., Kenyon, R.V. and Troxel, D.E. (1983). "Comparison of interpolating methods for image resampling". *IEEE Transactions on Medical Imaging*, vol. MI-2, no. 1, pp. 31-39, March.

[Peterson and Soderberg 89] Peterson, C. and Soderberg, B. (1989). "A new method for mapping optimization problems onto neural networks". *International Journal of Neural Systems*, 1(1): 3-22.

[Pratt 78] Pratt, W.K. (1978). *Digital Image Processing*, Wiley and Sons, New York, 2nd Edition.

[Qi et al. 97] Qi, H., Snyder, W.E. and Bilbro, G.L. (1997). "Using mean field annealing to solve anisotropic diffusion problems". *IEEE International Conference on Image Processing*, v3, pp. 352-355, Santa Barbara, CA.

[Qi et al. 98] Qi, H., Snyder, W.E. and Bilbro, G.L. (1998). "Missing data estimation by separable deblurring". *Proceedings for the IEEE International Joint Symposia on Intelligence and Systems*, pp. 348-353, Rockville, MD, May.

[Rabbani 88] Rabbani, M. (1988). "Bayesian filtering of Poisson noise using local statistics". *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36 (6): 933-937.

[Rosenfeld and Kak 82] Rosenfeld, A. and Kak, A.C. (1982). *Digital Picture Processing*, Academic Press, New York, vol. 1, 2nd Edition.

[Rothwell and Drachman 89] Rothwell, E. and Drachman, B. (1989). "A unified approach to solving ill-conditioned matrix problems". *International Journal for Numerical Methods in Engineering*, 28: 609-620.

[Schultz and Stevenson 94] Schultz, R.R. and Stevenson, R.L. (1994). "A Bayesian approach to image expansion for improved definition". *IEEE Transactions on Image Processing*, 3(3): 233-242, May.

[Snyder et al. 91] Snyder, W.E., Santago, P. Logenthiran, A., et al. (1991). "Segmentation of magnetic resonance images using mean field annealing". *XII International Conference on Information Processing in Medical Imaging*, Kent, England, July 7-11.

[Snyder 93] Snyder, W.E. (1993). *Industrial Robots: Computer Interfacing and Control.* Prentice-Hall.

[Snyder 99] Snyder, W.E. (1999). *Computer Vision Class Notes*, Fall Semester.

[Szeliski 89] Szeliski, R. (1989). *Bayesian Modeling of Uncertainty in Low-Level Vision.* Kluwer Academic Publishers, Boston.

[Tikhonov and Arsenin 77] Tikhonov, A.N., Arsenin, V.Y. (1977). *Solution of Ill-Posed Problems*, Winston, Wiley, New York.

[Trefethen and Bau 97] Trefethen, L.N., Bau D. III (1997). *Numerical Linear Algebra*, SIAM.

[Trucco and Verri 98] Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, New Jersy.

[Wang 96] Wang, C.X. (1996). *Optimal Image Interpolation Using Optimal Method*, Ph.D. Thesis, North Carolina State University, Raleigh, NC.

[Watkins et al. 93] Watkins, C.D., Sadun, A. and Marenka, S. (1993). *Modern Image Processing: Warping, Morphing, and Classical Techniques*, Academic Press Professional, Chapter 1.

7. Appendixes

7.1 Image Acquisition

The image acquisition module consists of three parts: (1) *the CCD head module*, that converts optical signals to electrical signals; (2) *the CCD driver assembly*, that provides the system control signals, the sampling mechanisms, and the signal amplification circuits; and (3) *the analog-to-digital converter* (ADC), that quantizes and resamples (optional) analog signals and converts them to digital ones.

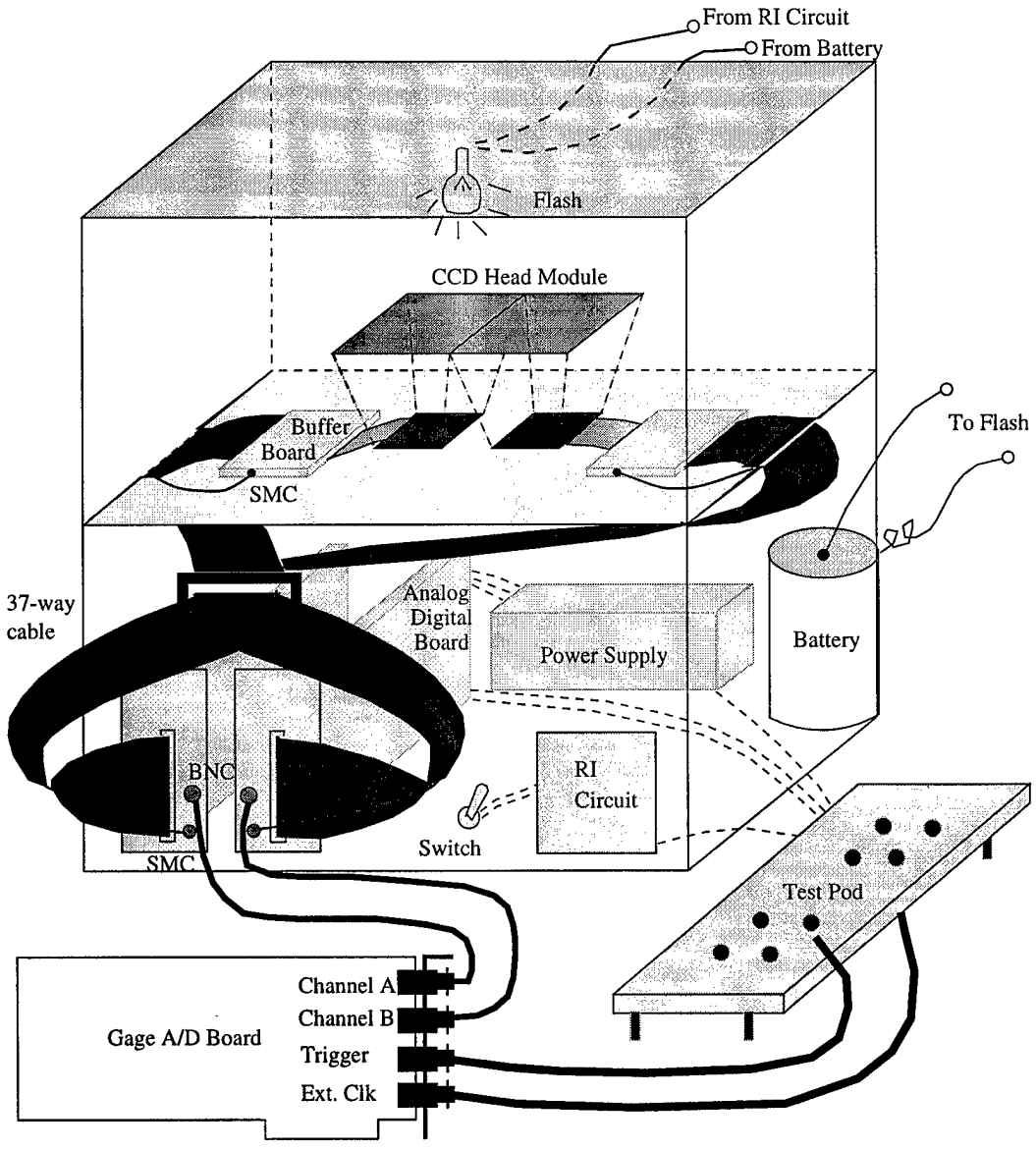


Figure 24. Front view of the image acquisition system assembly.

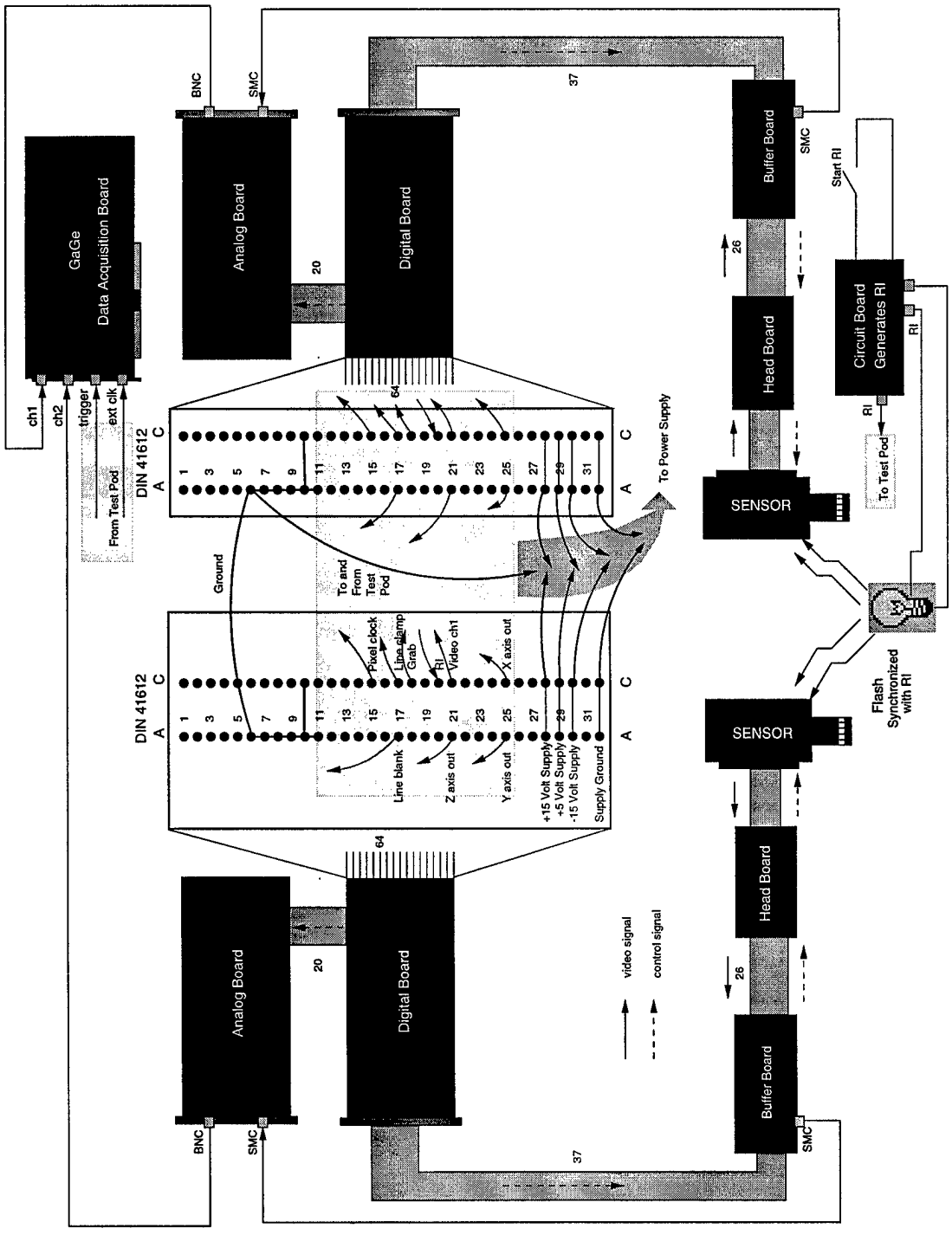


Figure 25. System connection diagram.

Fig. 24 shows a front view of the image acquisition system assembled in a Bud box. Flash is used as the light source instead of X-ray for experimental purpose. Fig. 25 is a simplified system hookup diagram.

7.1.1 CCD Head Module

The head module (from EEV, Inc.) performs like a detector and a transducer. It contains a scintillator screen, a 1 x 2 CCD sensor arrays, each of which is hard coupled to a fiber optic taper. The two tapers are dry cou-

pled together in a mechanical frame, so that the dead space between adjacent tapers is always less than 200μm, with 50μm the usual space.

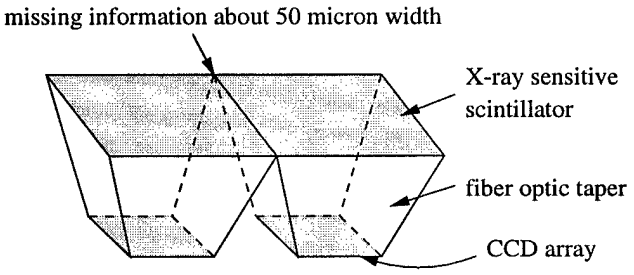


Figure 26. The CCD head module.

The scintillator uses a phosphor screen to convert an X-ray beam into myriad points of light. The phosphor screen is Eu doped Gd_2O_2S, 100μm thick.

Each fiber optic taper has a 50 x 50 mm^2 outer surface and a 25 x 25 mm^2 inner surface, which is usually less than the CCD active area. The CCD active area is 25.9 x 27.9 mm^2, consists of 1152 rows and 1242 columns, with each pixel size 22.5 x 22.5 μm^2.

Light is guided from the bottom of phosphor screen down to each CCD cell by a bundle of fibers. The CCD cell can then convert the light signal to an electrical signal by photoelectric process, and the generated charges are accumulated at each CCD cell.

Care must be taken not to allow too much light onto the CCD since a gross overload can cause the resulting image to turn black. Basically, too much light entering the CCD causes the video signal out of the device to look like there is no light entering the CCD (i.e. black). However, with no light entering the front of the optic taper but with light still able to enter through the sides, there is still enough light entering the CCD to cause the video signal out of the CCD to become saturated (i.e. white). Thus careful control of the lighting conditions is necessary to achieve correct results. In our experiments, the Bud box is covered by opaque black paper and the edges of the fiber optic tapers are covered by black insulating tape to prevent light leakage into the CCD. The flash light source is covered by two layers of black insulating tapes, with a 0.5 x 0.5 cm^2 square hole in the outer layer tape as our point source to reduce the amount of light getting to the CCD.

7.1.2 CCD Driver Assembly

CCD driver assembly (from EEV, Inc.) drives the CCD readout. Three boards are involved in this process: digital board, analog board, and buffer board (Fig. 25).

Digital Board. The digital board houses the control logic for the CCD sensor and the video processing. It is also responsible for interfacing with the outside world. The digital board sends control pulses to the analog board via a 20-way piggyback connector, and to the buffer board via a two-meter 37-way ribbon cable. The 64-way (a and c) DIN 41612 connector is where power is fed into the CCD Driver Assembly and where the control and synchronization signals enter and leave. The DIN connector we use is from Harting (0903 164 6821). It is a female connector with 13mm wrap posts, 64 contacts mounted on columns *a* and *c*.

Four of the most important control signals are *pixel clock*, *line blank*, *grab*, and *request integration* (RI), where the line blank serves as the *horizontal synchronization* signal, and the grab as the *vertical synchronization* signal. The timing diagram is shown in Fig. 27.

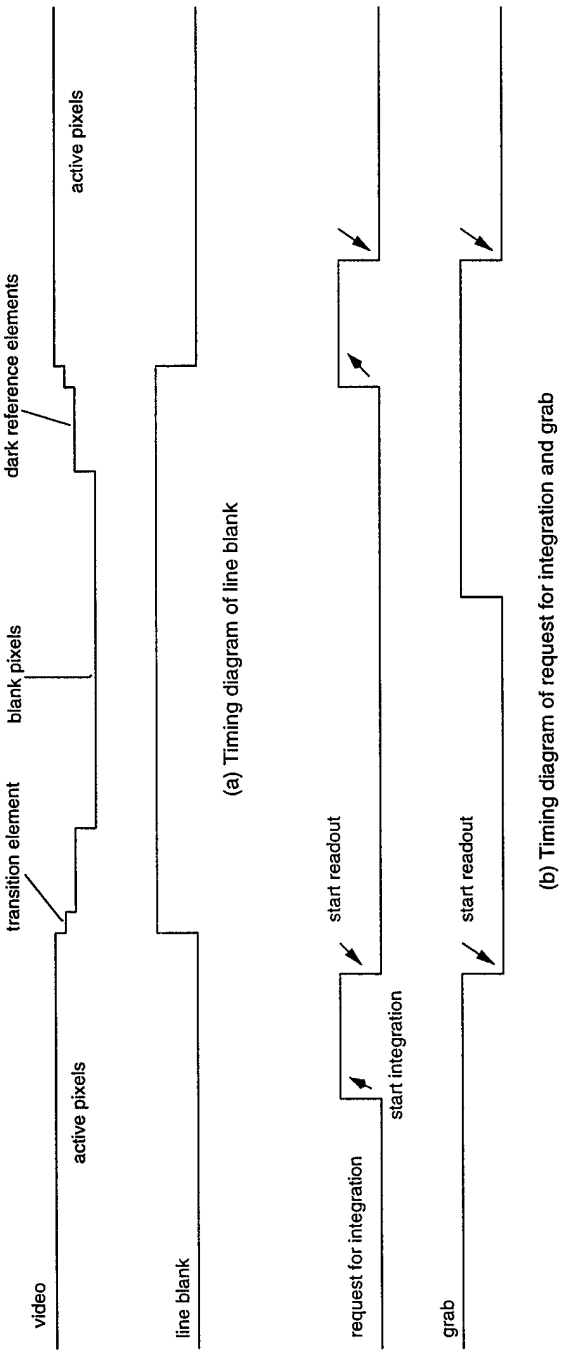


Figure 27. Timing diagrams of line blank, grab, and RI.

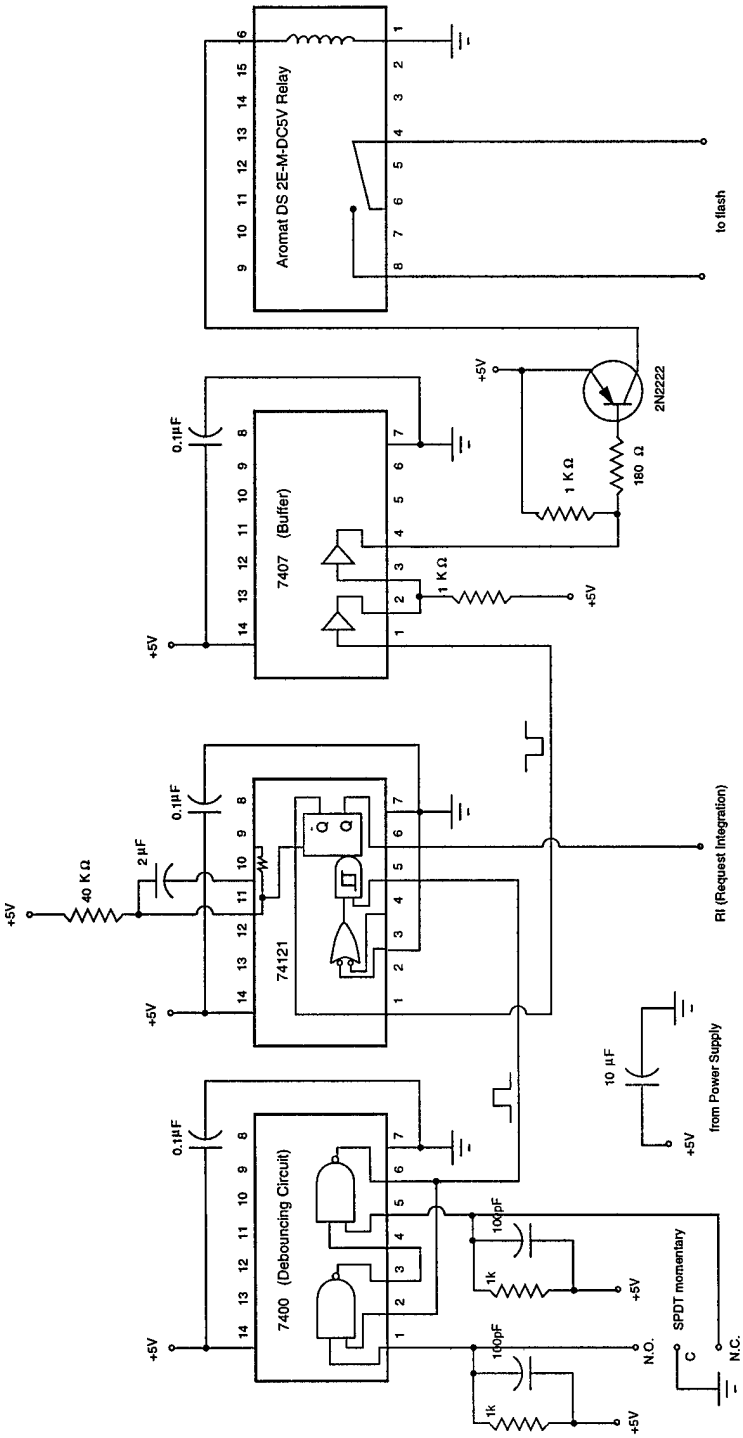


Figure 28. The circuit design for RI signal.

- Pixel clock is a 200ns pulse at a 1MHz rate. It occurs once for every pixel read out from the CCD. It is HIGH during the valid portion of each pixel in the video output from the analog board.

- Line blank is a pulse that occurs once every line read out from the CCD. It is LOW during the output of active video, HIGH during line transfer, readout of blank elements, dark reference elements and transition elements (Fig. 27 (a)).
- Grab is used to identify active video from the CCD. Grab is LOW during the readout of active lines, HIGH at all other times (Fig. 27 (b)).
- Request for integration (RI) determines the integration interval. When RI goes HIGH, clocking stops, integration starts; when it goes LOW, readout starts (Fig. 27 (b)). RI triggers the Grab signal. Fig. 28 is the circuit design for generating RI signal. The integration interval is about 100 ms.

Buffer Board and Analog Board. The buffer board terminates the control signals from the digital board to the sensor and amplifies the output signal from the sensor. The sensor signal is sent via a mini-COAX cable (SMC port) to the analog board (Fig. 24, Fig. 25).

The analog board processes the sensor signal from the buffer board using *correlated double sampling* (CDS) to isolate the signal level difference, and produce analog video, with adjustable gain. The video signal is linked to a BNC connector on the front panel which will go into the input channel of the A/D board from GaGe.

Amplification and Readout Settings. The signal read out from the CCD head module is very weak, and amplification must be conducted before digitization. In this driver assembly, gains are set at three stages of the video path. The first gain occurs on the buffer board, where an amplifier functions with adjustable gains at 4, 8, or 16. The second is on the analog board with adjustable gains at 1 or 4 that adjusts the input to the CDS circuit. The last gain stage is under control of a parameter switch on the digital board, that adjusts the output from the CDS circuit at x1, x2, or x4.

There are 6 switches on the digital board (Fig. 29), three of which are used to set up important parameters that control the CCD readout. Following is an overview of these three switches, and the appropriate settings for our system.

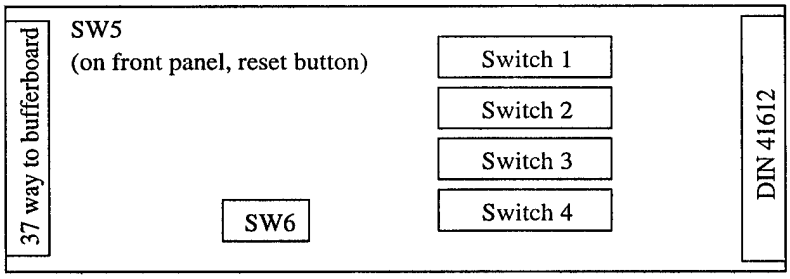


Figure 29. Switches on the digital board.

Switch 1 controls the readout mode, and the readout rate. Switch 2 sets up the gains for final video readout. Switch 5 is a front panel reset button, which resets the power supply of the driver board. Switch 6 indicates the model of the CCD assembly. Table 2 is the settings for each switch, and Table 3 is a brief explanation of the parameters that the switch controls. More details are addressed in [EEV 94].

TABLE 2. Switch settings.

Switch		a	b	c	d	e	f	g	h
Switch 1 00010101	on								
	off								
Switch 2 00011100	on								
	off								
Switch 3 00000000	on								
	off								
Switch 4 00000000	on								
	off								
Switch 6 0100	on					indicates model CCD05-30			
	off								

TABLE 3. Explanation of switch parameters.

	Switch 1		Switch 2	
	on	off	on	off
a	Frame transfer	Full frame	Direction of vertical readout charge (0: down; 1: up)	
b	Standard mode	Inverted mode		
c	Readout rate The setting for 1MHz is 010		Direction of horizontal readout charge (01: right; 10: left)	
d				
e			Gain setup for output of CDS circuit (00: x8, 10: x4, 01: x2, 00: x1)	
f	Normal	TDI mode		
g	Remote/Computer	Local Switch	N/A	
h	Visible light	X-ray source	N/A	

7.1.3 Other Supply Accessories

Besides the CCD driver assembly, other supply accessories are needed for successful CCD readout. These accessories include RI circuit board, power supply, test pod, etc.

Request integration (RI) is a key control signal in the driver assembly. When RI is high, the CCD sensor stops clocking, allowing light accumulation, which is the integration process. After a 100ms integration interval, RI goes LOW, and drives GRAB to go HIGH to start the readout process. RI circuit board (Fig. 28) is designed to generate a 100ms pulse, controlled by an SPDT momentary switch.

To synchronize the start of integration and the flash, the RI circuit board adopts some extra design ideas to achieve high-speed, noise-free logic [Snyder 93]. A large tantalum capacitor (10 to 20 μF) bypasses the board to suppress the low-frequency external noise. Each chip on board is bypassed with 0.001 to 0.01 μF ceramic or monolithic capacitors between Vcc and ground to avoid sudden voltage drop when chips demand sudden bursts of current.

To support the CCD head module and driver assembly, a power supply is required to provide +5 volts at 400mA, +15 volts at 500mA, and -15 volts at 300mA. For a 1 x 2 CCD sensor arrays, in order to drive both

of the CCD sensors at the same time, the current requirements for each voltage level are doubled. That is, +5 volts at 800mA, +15 volts at 1A, -15 volts at 600mA.

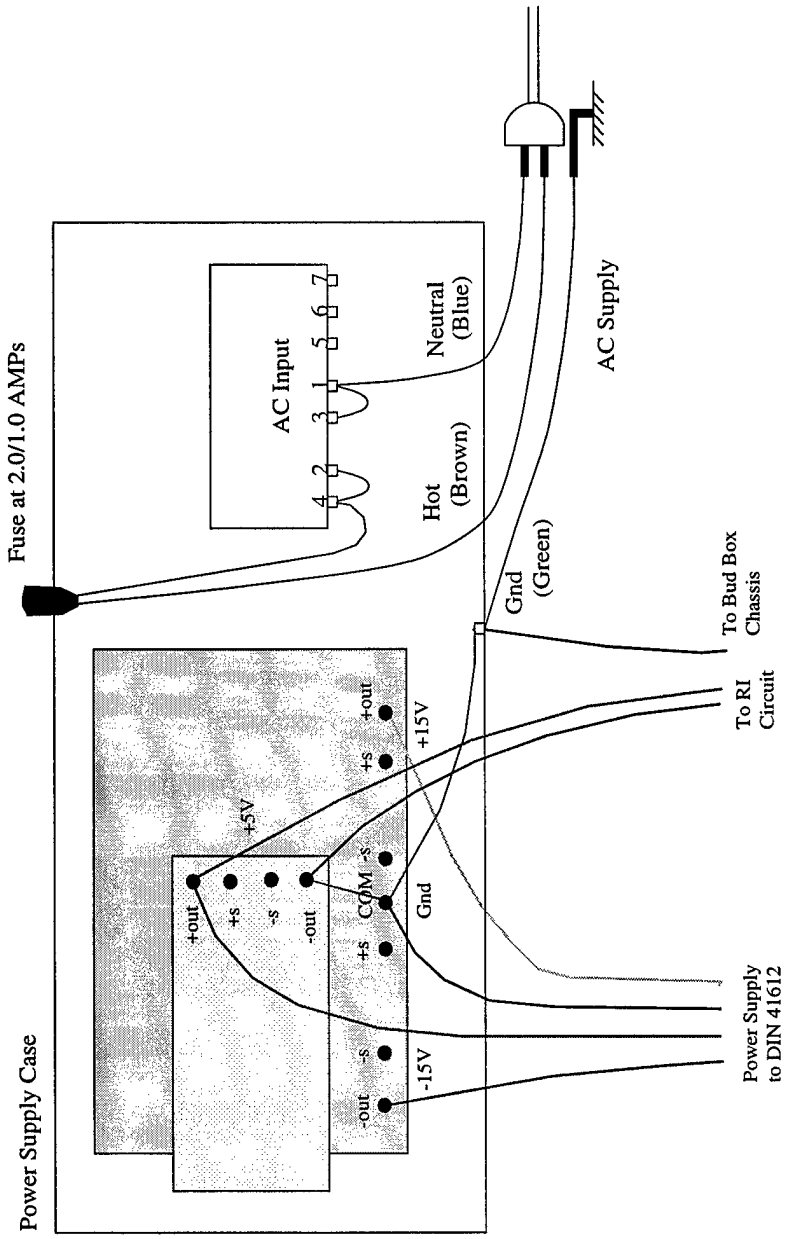


Figure 30. Power supply (International Power PC Power Supplies - IHBCC-512) connection diagram.

The power supply we used is from International Power DC Power Supplies. The IHBCC-512 model provides three outputs: 5V @ 3A, 15V @ 3A, and -15V @ 3A, which is enough for our application. Fig. 30 shows the connection diagram inside the power supply.

Test pod (Fig. 31) is used to test the signals from the DIN connector that are generated by the digital assembly. It is also used to switch between channel 1 control signals and channel 2 control signals to correctly trigger the ADC.

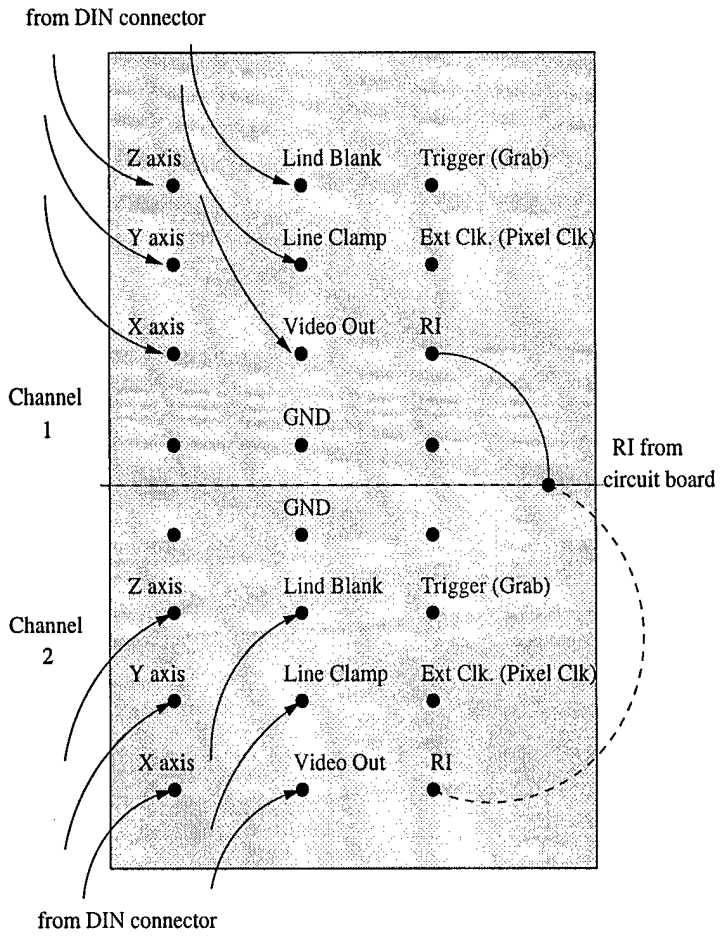


Figure 31. Test pod signals.

7.1.4 Analog-to-Digital Converter

The GaGe data acquisition board functions as a 12 bit A/D converter which can quantize and resample the analog signal into a 12-bit digital signal. The maximum data transfer rate is approximately 1.5 MBytes/sec.

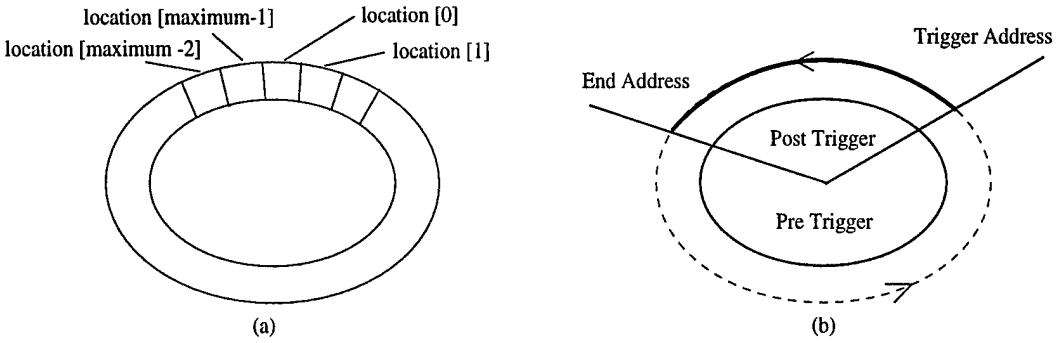


Figure 32. The circular on-board memory of the Gage data acquisition board.

The data coming out of the ADC is stored in the static memory (SRAM) of the GaGe board which is configured as a circular buffer (Fig. 32 (a)). Circular buffer is used to guarantee that the system will keep on capturing data indefinitely until a trigger event is detected.

Once initialized, the board starts to read in data, and at the same time waiting for trigger event to happen. The data that read in during this period is called *pre-trigger data.* When trigger event is received, a specified number of pixels are captured. These data are called *post-trigger data.* After storing the specified number of post-trigger data, the acquisition is stopped (Fig. 32 (b)).

The board we used is model CS512, which must be working under dual channel mode. That is, the two channels of data can be input at the same time. The array elements *location*[0] and *location*[1] hold the first pixel value from channel 1, with *location*[1] storing the least significant 8 bits, and *location*[0] storing the most significant 8 bits. Similarly, *location*[2] and *location*[3] hold the first pixel value from channel 2.

The sample rate of the ADC is set at 2MHz, with each channel at 1MHz, thus matching the CCD readout rate. Since the on-board memory of this model is 1M, with 512k for each channel, three integrations and readouts are required to capture images from one CCD sensor array, and six from both arrays.

With a 1MHz readout rate, 1M sample rate, 1152 vertical pixels and 1242 horizontal pixels (equating to 1,430,784 total pixels in an array), the performance of our data acquisition module can be summarized as follows: 1.43s minimum is required to readout a complete frame; 1.24ms minimum is required to readout a complete line; and 1μs is required to readout 1 pixel of information.

7.2 Image Display

Image display module may be needed anywhere, from measured image display, corrected image display, to restored image display.

In our system, the spatial resolution of a complete image is roughly 1152 x 2484, involving a 1 x 2 1152 x 1242 arrays. Since most of the current commercial monitors can only display image with dimensions 1024 x 1024 or smaller, how to efficiently and conveniently display larger images on smaller display planes is another problem which needs to be solved.

Here, we first discuss two possible approaches for image display: (1) use the *on-board video card*; and (2) use *a secondary display board.* We then propose to combine these two approaches and achieve *dual-screen mode loupe display* in our system.

7.2.1 Image Display through On-Board Video Card

This is the most common approach for image display. Since our A/D board works only under Windows operating system, we choose Visual C++ as the programming language, and use VisSDK (Vision Software Developers Kit) as the display library. VisSDK is a low-level C++ library, developed by Microsoft Vision Group. It aims to provide a strong programming foundation for research and application development, and includes possible real-time imaging. VisSDK is organized into different projects and VisDisplay is the one covers display related functions.

There are two ways to display a single image in VisDisplay. We can either get access to the raw pixel data and build a Windows bitmap from it, or we can use a Windows HDC (handle of a device context). A device

context (DC) is essentially a unified interface to a specific graphical device. It can be used to obtain information about the device and also perform graphical output to the device.

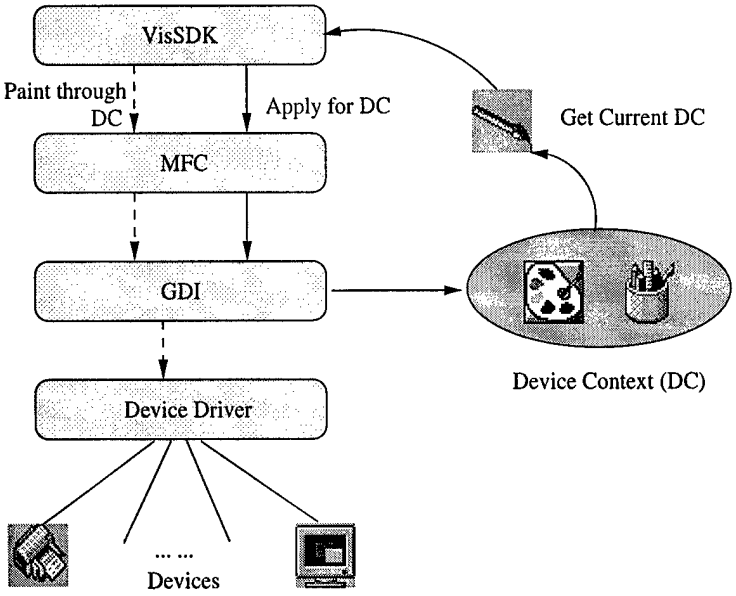


Figure 33. Software layers in image display.

Fig. 33 is a software layer diagram. At the lower level, device-dependent drivers are used to operate on the corresponding hardware directly. At a higher level, the graphics device interface (GDI) performs primitive device-independent graphical functions through DC. MFC (Microsoft Fundamental Class) is another layer that is built upon GDI. It consists of a series of predefined classes which save the users from tedious primitive GDI function calls, and provide a more convenient interface.

Once the image is read into the host memory, it can be displayed in a window's client area. Scrolling in both horizontal and vertical directions is needed for large image display. A window's client area is a versatile surface that can display anything a Windows program can draw.

7.2.2 Image Display by Secondary Display Board

The secondary display board we used is the Matrox Pulsar from Matrox Electronic Systems Ltd. This board features both on-board grab and display capabilities. The reason we did not use its grab functionality is because the minimum pixel clock it can accept is 2MHz, while the pixel clock from our image acquisition system is 1MHz. However, the Matrox board provides a powerful display section, featuring a 2M image frame buffer, a 2M graphics overlay (VGA) frame buffer for non-destructive overlay capabilities, and dual-screen mode display.

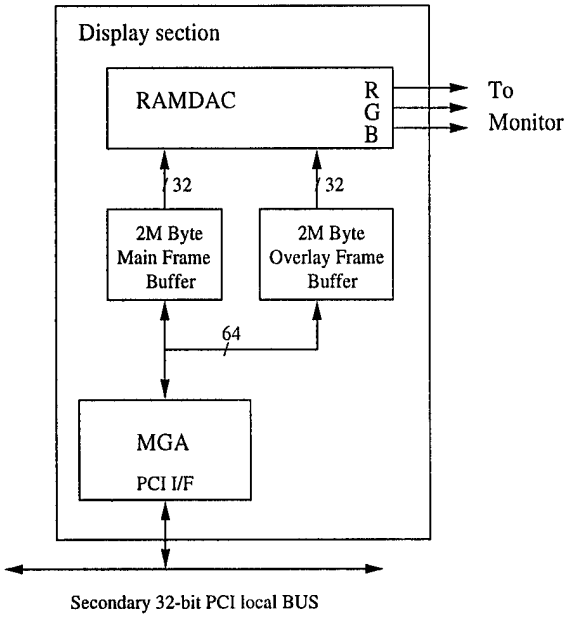


Figure 34. Display section of Matrox Pulsar.

The display section (Fig. 34) is also powered by the Matrox MGA-2064W graphics accelerator that can display at resolutions up to 1600x1200x256 colors. This graphics accelerator can be used either as the main display controller of the host system; or as a separate graphics board so that users can work in dual-screen mode. The overlay (VGA) and main image frame buffers support common zooming (by factors of 2 or 4), panning, and scrolling.

A software package called MIL (Matrox Imaging Library), developed by Matrox, provides interfacing function calls to the display board.

The big advantage of using a second display board is the dual-screen display mode. The users can display images on one monitor, while doing controls on another.

7.2.3 Dual-Screen Mode Loupe Display

In medical imaging, very often some detail analysis is needed on a few interesting sections of the image, such as a suspicious lesion area. For the purpose of conveniently processing sections of images, while at the same time keeping a complete viewgraph of the original one, we propose to combine the use of both approaches we discussed above, and implement the so-called *dual-screen model loupe display.*

We use the word *loupe* because it is analogous to the operation of zooming a patch of the original image. To keep viewgraphs of images at different scales, dual-screen is a natural choice. The Matrox Pulsar board also provides built-in functions for zooming and color mapping, that can be easily used to operate on the zoomed patches of the original image.

Object-oriented programming language is used in developing this technique. Three objects are extracted: the *Gage* ADC, the *Pulsar* display board, and the *image*. *Gage* object consists of methods for acquiring data from the CCD sensor arrays. *Image* object involves methods for image display on the host monitor, and for

image segmentation. *Pulsar* object is responsible for displaying segments of images onto the secondary monitor, along with methods for zooming, color mapping, and contrast tuning, etc. (Fig. 35).

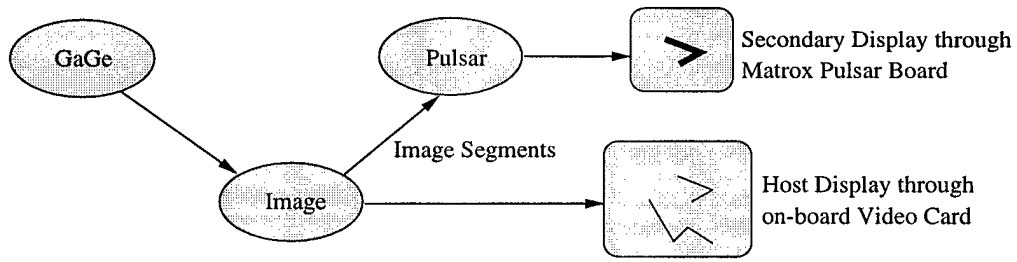


Figure 35. Object-oriented display design.

We developed our display system on Windows NT platform. Images are captured using the interfacing functions provided by GaGe ADC. The complete image is displayed on the main monitor (host display). The secondary monitor is used to display regions of interest based upon user's selection. Users can either mark the region of interest by a rectangular, or they can click on any interesting pixel with a surrounding region of 64 x 64 or 128 x 128 automatically extracted from the original image.

The region of interest is displayed on the secondary monitor, where several image enhancement techniques can be operated. These techniques include zooming at any integer scale, autoscaling, contrast stretching, pseudo-color enhancement (use color to help visual discrimination of small grey level differences), edge enhancement (convolve image with high-pass filter kernel such as Laplacian type kernels), unsharp masking (equivalent to high-pass spatial filtering), etc.

7.3 Sources of Radiometric and Geometric Distortions

In our system, the functions of the fiber-optic tapers are three-fold: (1) used as an aligner to make adjacent CCD detectors aligned very close to each other such that only 1~2 columns of information is missed along the butting edge; (2) used as the light-guide system, where a bundle of fibers connect each point on the scintillator screen to a corresponding cell on the CCD detector; and (3) used as a demagnifier to increase the field of view of the CCDs, which is usually in the ratio of 2:1.

However, the use of the fiber-optic tapers also causes radiometric and geometric distortions, which originates during manufacture. Fig. 36 illustrates how a fiber-optic taper is constructed. A fiber-optic tube is first heated from the middle; then compressed to the middle, such that the area of the cutting surface matches the size of a CCD array to which it will attach; finally a cut from the middle makes two fiber-optic tapers with the smaller end attaches to the CCD array and the other end the outer-surface, facing the scintillator.

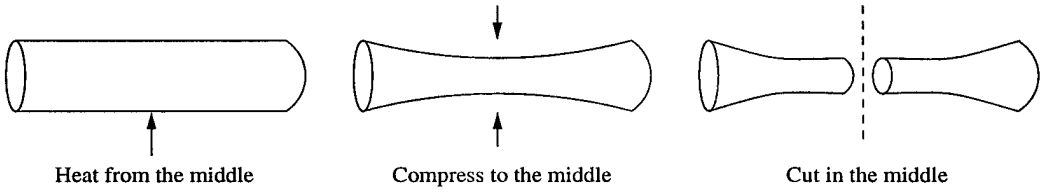


Figure 36. The manufacturing of the fiber-optic tapers.

In this process, we can see that no matter how precise the compression process, the cutting surface of the taper can not perfectly match the size of the CCD array. Usually, the taper is more compressed than it should be in order to avoid missing data sensed at the boundary of the outer-surface of the taper. That is why the captured image is always less than the CCD active area. The compression thus causes the "pincushion distortion", a well-known type of geometric distortion, where both the horizontal and vertical lines bend inwards toward the center of the display (Fig. 38). Instead of each bundle of fiber being guided to a corre-

sponding cell of the CCD detector, the bundles at the boundaries of the taper is deformed like Fig. 37. The distortion is not symmetric, exhibiting some degree of shear effect.

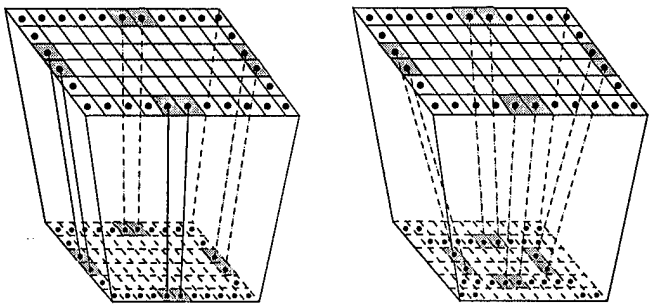


Figure 37. The correct alignment (left), the misalignment of fiber optic bundles and CCD cells along the edge of the detector which causes the pincushion distortion (right).

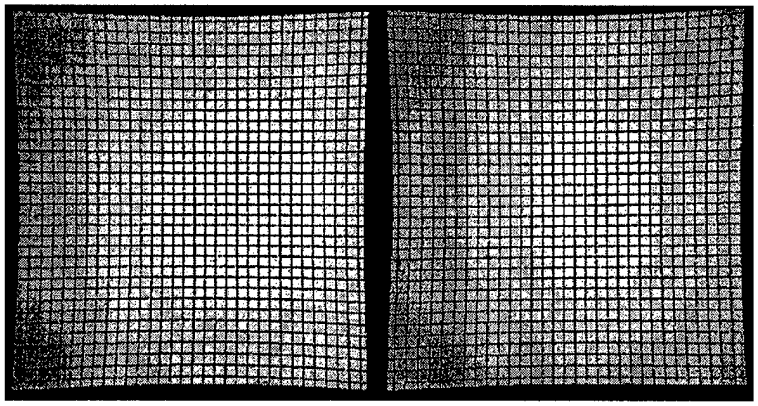


Figure 38. Vignetting-type radiometric distortion and geometric distortion.

Furthermore, because of the shape of the taper, the fibers at the boundary need to travel a longer distance than those at the center, and the light transfer is less efficient when the fiber is not normal to the CCD surface. Thus, there is an intensity gradation from the center to margin like a cross section of a tree, or a set of concentric ellipses (Fig. 38). This kind of distortion is called the *vignetting-type* radiometric distortion, as it resembles the vignetting which results from an imperfect lens.

7.4 Radiometric Correction

Radiometry is the study of image formation. It concerns two sources of light energy [Trucco and Verri 98]: the amount of light reflected from the object, and the amount of reflected light that actually reaches the sensor. Assume the amount of light reflected from the object is homogeneous; if there is no radiometric distortion, the captured image should have homogeneous illumination over the whole image. However, as mentioned in Appendix 7.3, the vignetting-type radiometric distortion corrupts the homogeneity.

The fundamental equation of radiometric distortion is Eq. (4) [Trucco and Verri 98], where $g(x, y)$ and $f(\xi, \eta)$ are the *image irradiance* (the power of the light at each point of the image plane per unit area) and the *image radiance* (the power of the light ideally emitted by each point of a surface in 3-D space in a given direction) respectively. Eq. (4) says that the image irradiance, $g(x, y)$, decreases as the fourth power of the cosine of the

angle (α) formed by the principal axis with the optical image irradiance (Fig. 39). $g(x, y)$ is also regarded as uniformly proportional to the scene radiance, $f(\xi, \eta)$, over the whole image plane.

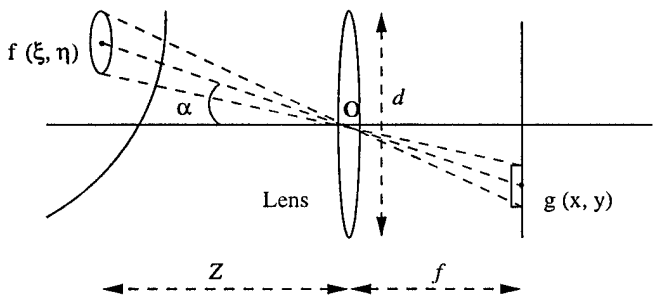


Figure 39. Radiometry in image formation.

$$g(x, y) = f(\xi, \eta)\frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4\alpha \propto f(\xi, \eta)\cos^4\alpha \tag{4}$$

Conventional method [Frei 83] simplifies the radiometric model of Eq. (4) by a *polynomial approximation* as Eq. (5),

$$g(x, y) \approx \frac{f(x, y)}{a_0 + a_1(x^2 + y^2)} \tag{5}$$

We observe from Fig. 38 that the illumination distribution of both the left and right part of the image shaped like ellipses with different length of axes. Therefore, we modify Eq. (5) into Eq. (6), where the equation for an ellipse is used to model the distortion. (x_i, y_i) are the coordinates of pixel i, and (x_o, y_o) denotes the origin of the concentric ellipses.

$$g(x_i, y_i) \approx \frac{f(x_i, y_i)}{a_0 + a_x(x_i - x_o)^2 + a_y(y_i - y_o)^2} \tag{6}$$

The coefficients $a = (a_0, a_x, a_y)$ must be determined in order to be used to correct the distorted image g back into $\hat{f}$ (Eq. (7)), such that Eq. (8) can reach its minimum. M is the total number of pixels in the image.

$$\hat{f}_i = g(x_i, y_i)[a_0 + a_x(x_i - x_o)^2 + a_y(y_i - y_o)^2] \tag{7}$$

$$\varepsilon = min_a \sum_{i=0}^{M-1} (\hat{f}_i - f_i)^2 \tag{8}$$

To make the computation easier, we rewrite Eq. (7) and Eq. (8) using matrix notation. Let

$$W = \begin{bmatrix} 1 & (x_0 - x_o)^2 & (y_0 - y_o)^2 \\ 1 & (x_1 - x_o)^2 & (y_1 - y_o)^2 \\ \cdots & \cdots & \cdots \\ 1 & (x_{M-1} - x_o)^2 & (y_{M-1} - y_o)^2 \end{bmatrix}$$

$$A = \begin{bmatrix} a_0 \\ a_x \\ a_y \end{bmatrix}$$

$$G = \begin{bmatrix} g_0 & 0 & 0 & 0 & \dots & 0 \\ 0 & g_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & g_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & g_{M-1} \end{bmatrix}$$

$$\hat{F} = \left[\hat{f}_0 \cdots \hat{f}_{M-1}\right]^T$$

then Eq. (7) can be written as

$$\hat{F} = GWA$$

and Eq. (8) as Eq. (9),

$$\varepsilon = min_A (F - GWA)^T (F - GWA) \tag{9}$$

The solution to Eq. (9) is Eq. (10),

$$A = (GW)^{-1} F \tag{10}$$

where $(GW)^{-1}$ is the pseudo-inverse, which can be computed by

$$(GW)^{-1} = [(GW)^T (GW)]^{-1} (GW)^T$$

In real world applications, we do not solve Eq. (9) based on the whole set of image pixels, which would be very time consuming and unnecessary. Instead, we choose a set of m pixels as the *control points*. The control points are the ones that we know what their original intensity values should be.

We choose a rectangular grid as the testing image (Fig. 38). If not for the radiometric distortion, blur or noise, the measured image would be binary, with only two intensity levels: 255 and 0. We neglect noise for the time being, and choose the center point of each square as the control points to avoid blur effects. The original intensity value of these control points should be equal to the intensity value of the origin of the ellipse.

With the control points chosen, we can easily construct matrices W, G, and F, and apply them to Eq. (10) to solve the coefficient matrix A. With the coefficients solved, a look-up table can be generated with each entry indicating the scale that should be imposed onto each pixel of the measured image $g(x, y)$ to correct the radiometric distortion.

7.5 Geometric Correction

In geometric correction, a transformation function is designed to map the control points from a known pattern (such as a regular grid) to their measured positions. The geometric distortion in any captured images can then be corrected by back-projection. Control points are points whose positions in the original scene are known *a-priori*. Polynomial approximation, usually in the 2nd degree, is the conventional transformation function most often used [Pratt 78][Rosenfeld and Kak 82].

In our imaging system, we implemented the polynomial method in a higher degree. We also adopted the thin-plate spline (TPS) as another type of transformation function. TPS is a classical method for interpola-

tion. Bookstein [Bookstein 89] first used TPS to model the deformations between two sets of landmark points. As far as we know, this thesis is the first application of the method to geometric correction.

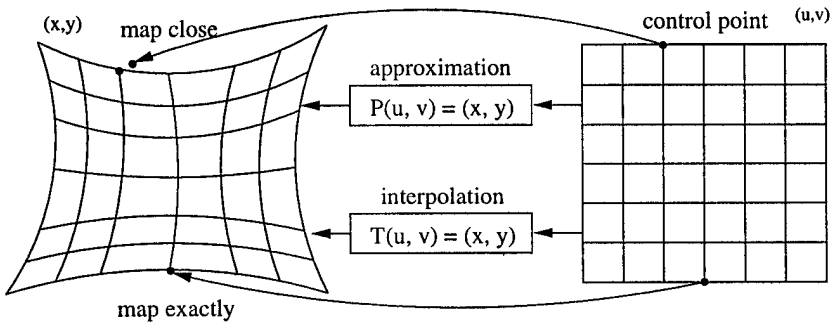


Figure 40. The transformation system between the measured image and the corrected image.

The difference between *approximation* and *interpolation* lies in the fact that approximation methods generate transformations that map all the control points *close* to their correspondence, so that the summation of displacements achieves a global minimum; whereas interpolation methods produce transformations where all the control points can be mapped to their correspondence *exactly* [Daview and Samuels 96].

7.5.1 Polynomial Approximation

Nearly all the methods used to correct pincushion distortion so far are variants of polynomial approximation [Beier et al. 92][Butler and Pierson 91]. Pratt addressed the 2nd degree polynomial approximation in 1978 [Pratt 78]. Here, we generalize this method to an nth degree polynomial.

Assume $\left[\hat{x}_i\ \hat{y}_i\right]^T = \left[P_x(u_i, v_i)\ P_y(u_i, v_i)\right]^T$, where (u_i, v_i) is a control point from the corrected image, $(\hat{x}_i, \hat{y}_i)$ is the corresponding point in the measured image, and (P_x, P_y) are the transformation functions for the x and y coordinates respectively, both of which are nth degree polynomials with respect to u and v. P_x and P_y are expressed in Eq. (11) and Eq. (12),

$$\hat{x} = P_x(u, v) = \sum_{i=0}^{n-1} \sum_{r+s=i} a_{irs} u^r v^s \tag{11}$$

$$\hat{y} = P_y(u, v) = \sum_{i=0}^{n-1} \sum_{r+s=i} b_{irs} u^r v^s \tag{12}$$

where a_{irs} and b_{irs} are the coefficients. The transformation functions should map all control points (u, v) to (x, y) as closely as possible, that is, $(\hat{x}, \hat{y})$ as close to (x, y) as possible. Minimum mean square error (MMSE) is used to measure the error as Eq. (13),

$$\varepsilon = min_{a,b} \sum_{i=0}^{m-1} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \tag{13}$$

where m is the number of control points. Coefficients a and b are determined so that ε can reach its minimum.

Since the pincushion distortion that our imaging system suffers is not symmetric, we use a 3rd degree polynomial to model the transformation. Based on Eq. (11) and Eq. (12), we get

$$\hat{x} = a_{000} + a_{110}u + a_{101}v + a_{220}u^2 + a_{211}uv + a_{202}v^2 + a_{330}u^3 + a_{321}u^2v + a_{312}uv^2 + a_{303}v^3$$

and

$$\hat{y} = b_{000} + b_{110}u + b_{101}v + b_{220}u^2 + b_{211}uv + b_{202}v^2 + b_{330}u^3 + b_{321}u^2v + b_{312}uv^2 + b_{303}v^3$$

Again, the transformations are reformulated in matrix form. Let

$$A = \left[a_{000}\ a_{110}\ a_{101}\ \cdots\ a_{303}\right]^T,\ B = \left[b_{000}\ b_{110}\ b_{101}\ \cdots\ b_{303}\right]^T$$

$$X = \left[x_0\ \cdots\ x_{m-1}\right]^T,\ Y = \left[y_0\ \cdots\ y_{m-1}\right]^T$$

$$\hat{X} = \left[\hat{x}_0\ \cdots\ \hat{x}_{m-1}\right]^T,\ \hat{Y} = \left[\hat{y}_0\ \cdots\ \hat{y}_{m-1}\right]^T$$

$$W = \begin{bmatrix} 1 & u_0 & v_0 & u_0^2 & u_0v_0 & v_0^2 & u_0^3 & u_0^2v_0 & u_0v_0^2 & v_0^3 \\ 1 & u_1 & v_1 & u_1^2 & u_1v_1 & v_1^2 & u_1^3 & u_1^2v_1 & u_1v_1^2 & v_1^3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_{m-1} & v_{m-1} & u_{m-1}^2 & u_{m-1}v_{m-1} & v_{m-1}^2 & u_{m-1}^3 & u_{m-1}^2v_{m-1} & u_{m-1}v_{m-1}^2 & v_{m-1}^3 \end{bmatrix}$$

such that,

$$\hat{X} = WA, \quad \hat{Y} = WB$$

and

$$\varepsilon = min_{A,B}[(X - WA)^T(X - WA) + (Y - WB)^T(Y - WB)] \tag{14}$$

The solution to Eq. (14) is

$$A = W^{-1}X, \quad B = W^{-1}Y$$

where W^{-1} is the pseudo-inverse, computed by $(W^TW)^{-1}W^T$.

7.5.2 Thin-Plate Spline Interpolation

Interpolation differs from approximation in the way that interpolation problem tries to construct a continuous function $f: \Re^n \rightarrow \Re^m$ so that $f(p_i) = q_i$, where p_i and q_i are finite sets of distinct point of $\Re^n$ and $\Re^m$ respectively; while approximation can only find function $g: \Re^n \rightarrow \Re^m$ such that $g(p_i) \approx q_i$. In order to achieve *well-posed* (existence, uniqueness, and stability) solution to the interpolation problem, it proves quite natural to require the minimization of some appropriate expression, such as the quadratic seminorm[1] $|f|^2$ [Meinguet 84]. $|f|^2$ can be physically interpreted as the *bending energy of a thin plate of infinite extent.* Therefore, interpolants of minimum seminorm can be appropriately termed *thin plate splines* (TPS).

1. A real-valued functional $|x|$, defined on a vector space X, is called a seminorm if: a) $|x| \geq 0$ for all $x \in X$; b) $|\alpha x| = |\alpha| \cdot |x|$; c) $|x + y| \leq |x| + |y|$

Bookstein [Bookstein 78][Bookstein 89] is the first researcher to use TPS to model the deformations between two sets of landmark points (also called control points), where TPS warps one picture into the space of another by mapping landmarks exactly onto landmarks while minimizing a plausible *bending energy*, the integral of summed squared second derivatives of the mapping as Eq. (15). In the case that $f(x, y)$ is a brightness function, the integrand is referred to as the *quadratic variation*.

$$|f|^2 = \iint_R \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy \tag{15}$$

In the special case of geometric correction, TPS represents the mapping of control points from the corrected image to their correspondence in the measured image in the manner which *uniquely* minimizes the *bending energy* of Eq. (15).

In Eq. (15), $\Delta^2 f = (\partial^2 f/\partial x^2)^2 + 2(\partial^2 f/\partial x \partial y)^2 + (\partial^2 f/\partial y^2)^2$ is the *biharmonic equation*, and the fundamental solution of $\Delta^2 f = 0$ is $U(r) = r^2 \log r^2$, where r is the distance between two control points. Since $\Delta^2 f$ is non-negative, the unique solution that minimizes the bending energy is a linear combination of $U(r)$, which is bounded and asymptotically flat. If scaling or homogeneous shear is allowed before applying the displacements, an affine transformation should be added to the solution to indicate the behavior of f at infinity. Therefore, the complete solution to Eq. (15) is Eq. (16):

$$f(x, y) = \sum_{j=1}^{m} w_j U(r_j) + a_1 + a_x x + a_y y \tag{16}$$

where m is the number of control points. Eq. (16) consists of two parts: the *non-linear* part which is the sum of functions $U(r)$, and the *linear* part which is the affine transformation.

In our imaging system, we adopt TPS to correct geometric distortion. We still assume (x_i, y_i) to be a set of points from the measured image, and (u_i, v_i) from the corrected image. The TPS transformation function T: $R^2 \rightarrow R^2$ is sought such that

$$\left[x_i \; y_i \right]^T = \left[T_x(u_i, v_i) \; T_y(u_i, v_i) \right]^T \tag{17}$$

and T minimizes the bending energy of Eq. (15). Substituting Eq. (16) into Eq. (17), we get

$$\begin{bmatrix} \sum_{j=0}^{m-1} p_i U(r_{ij}) + a_1 + a_x u_i + a_y v_i \\ \sum_{j=0}^{m-1} q_i U(r_{ij}) + b_1 + b_x u_i + b_y v_i \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

where $r_{ij} = |(u_i, v_i) - (u_j, v_j)|$. Again, we rewrite the problem in matrix form as Eq. (18),

$$\begin{bmatrix} U(r_{0,0}) & U(r_{0,1}) & \cdots & U(r_{0,m-1}) & 1 & u_0 & v_0 \\ U(r_{1,0}) & U(r_{1,1}) & \cdots & U(r_{1,m-1}) & 1 & u_1 & v_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ U(r_{m-1,0}) & U(r_{m-1,1}) & \cdots & U(r_{m-1,m-1}) & 1 & u_{m-1} & v_{m-1} \end{bmatrix} \begin{bmatrix} p_0 & q_0 \\ p_1 & q_1 \\ \cdots & \cdots \\ p_{m-1} & q_{m-1} \\ a_1 & b_1 \\ a_x & b_x \\ a_y & b_y \end{bmatrix} = \begin{bmatrix} x_0 & y_0 \\ x_1 & y_1 \\ \cdots & \cdots \\ x_{m-1} & y_{m-1} \end{bmatrix} \tag{18}$$

which we denote as Eq. (19),

$$\begin{bmatrix} K_{m\times m} & L_{m\times 3} \end{bmatrix} \begin{bmatrix} P_{m\times 1} & Q_{m\times 1} \\ A_{3\times 1} & B_{3\times 1} \end{bmatrix} = \begin{bmatrix} X_{m\times 1} & Y_{m\times 1} \end{bmatrix} \tag{19}$$

To solve the coefficient matrices P, Q, A, and B, matrix $\begin{bmatrix} K & L \end{bmatrix}$ should be inverted. To make the matrix size compatible for inverting, Eq. (19) is reconstructed into Eq. (20), where O is the zero matrix with appropriate dimensions. The coefficients can then be solved by Eq. (21).

$$\begin{bmatrix} K & L \\ L^T & O \end{bmatrix} \begin{bmatrix} P & Q \\ A & B \end{bmatrix} = \begin{bmatrix} X & Y \\ O & O \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} P & Q \\ A & B \end{bmatrix} = \begin{bmatrix} K & L \\ L^T & O \end{bmatrix}^{-1} \begin{bmatrix} X & Y \\ O & O \end{bmatrix} \tag{21}$$

Similar to the radiometric correction: with the control points chosen, we can construct matrices K, L, X, and Y, and apply them to Eq. (21) to solve for the coefficient matrices P, Q, A, and B. With the coefficients determined, and the corresponding control points, a look-up table can be calculated with each entry indicating the coordinates that a pixel point in the corrected image should be looked up into from the distorted image.

The more control points chosen, the longer computation time needed. However, since this kind of computation only happens once, and the mapping can be saved in a lookup table afterwards, the efficiency should not be a concern.

7.6 Degradation Analysis

[Aghdasi et al. 94] has given a detailed analysis of the degradations in an X-ray (film) image formation system. The complete model is illustrated in Fig. 41.

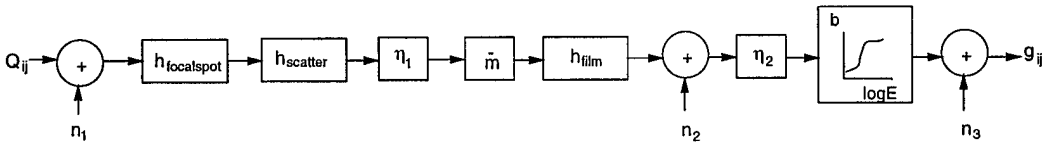


Figure 41. A model of radiographic image formation from [Aghdasi et al. 94].

The input (Q_{ij}) is the number of X-ray quanta received at the screen per pixel, and the output (g_{ij}) is the observed optical density of each pixel. This model consists of three sources of blur ($h_{focalspot}$, $h_{scatter}$, h_{film}), three scaling factors ($\overline{m}$, η_1, η_2), and three sources of noise (n_1, n_2, n_3).

The cause of $h_{focalspot}$ is that the focal spot on the anode of the X-ray tube is not a perfect point source, instead, it has a finite size, which generates the blur effect as shown in Fig. 42. $h_{scatter}$ results from the X-ray scatter in the object, and h_{film} from the film.

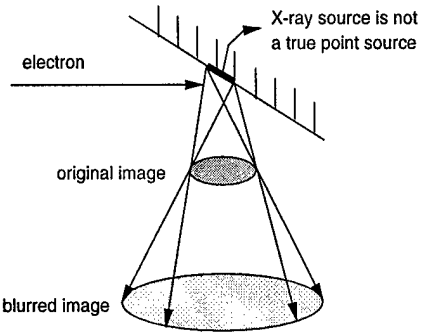


Figure 42. Blur caused by finite-size X-ray source.

Of the three scaling factors, $\overline{m}$ is the mean screen amplification ratio, η_1 and η_2 are the absorption efficiencies of the screen and the film respectively.

Noises mainly come from the film grain noise (n_3) and the quantum noise due to the discrete nature of the X-ray photons. Quantum noise can be further divided into the correlated components of the input quantum noise (n_1), and the uncorrelated components of the input quantum noise (n_2).

Our imaging system is different from Aghdasi's since we use CCD detectors instead of film. This has avoided all the degradations caused by film, but at the same time brought new types of degradations. For example, the butted fiber optic tapers introduce missing data along the butting edge. Therefore, a modified degradation model need to be formulated.

Blur. In our CCD imaging system, the blur caused by h_{film} is avoided. We also assume that the X-ray scatter can be largely absorbed by a vibrating grid. Therefore, the major blur source comes from the finite-sized X-ray source. A simple computation shows the effect of blur from this source:

In a digital mammographic system, it is reasonable to assume that the point source is 0.6mm width, the average distance from an object to the CCD detector is 2cm, and the distance from X-ray source to the object is 64cm, then the captured image of the focal spot will be 18μm. That is, the blur is about one pixel.

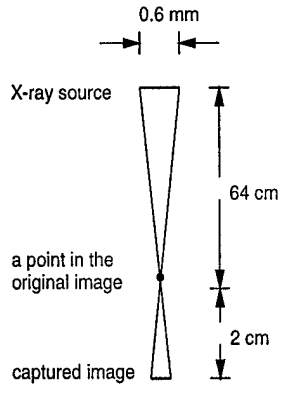


Figure 43. An example of blur effect by finite size focal spot.

Noise. Noise mainly comes from three sources in our system: 1) dark current noise which can be reduced by cooling; 2) quantum noise, a signal-depend noise which can be modeled by a Poisson process [Rabbani 88]. When the readout rate is very high and the amount of the readout data is very large, this kind of noise can be neglected; 3) fixed noise, due to the electrical system noise, quantization noise, etc., which is the major noise source in our system.

CCD Detector Induced Errors. If a single CCD cell is defective, it can cause data from the rest of the row missing due to CCDs serial readout mechanism. Also, butting technique can cause data at the boundary of adjacent CCDs missing. Both of these sources introduce missing column(s)/row(s) in images. Defects in the metallization or fibers can introduce single point defects as well.

7.7 Regularization Theory

Missing data estimation is basically an image restoration problem. Image restoration is well known to be *ill-posed* [Andrews and Hunt 77][Katsaggelos 91], since a little change in the input data can dramatically affect the solution. Ill-posedness comes from the *ill-conditioning* of the matrix constructed from the system PSF. The most popular solution to ill-posedness is *regularization.* Although many different kinds of image restoration algorithms have been proposed so far, they all share a common structure: the *regularization theory.*

Generally speaking, any regularization method tries to analyze a related *well-posed* problem whose solution approximates the original ill-posed problem. The well-posedness is achieved by implementing one or more of the following basic ideas [Nashed 76]:

- change of the concept of a solution;
- restriction of the data;
- change of the space and/or topologies;
- modification of the operator itself;
- the concept of regularization operators; and
- well-posed stochastic extensions of ill-posed problems.

According to [Tikhonov and Arsenin 77], a regularization method consists of finding regularizing operators that operate on the data, and determining the regularization parameter from supplementary information pertaining to the problem. The regularization operator depends continuously on the data and results in the true solution when the regularization parameter goes to zero, or equivalently when the noise goes to zero.

In the traditional image restoration approaches, image formation is modeled as Eq. (22),

$$Hf + n = g \tag{22}$$

where f, n, and g are $N^2 \times 1$ vectors, the lexicographic representation of the original image, noise, and the measured image respectively; and H is the $N^2 \times N^2$ Toeplitz matrix, derived from the system PSF. The regularization method constructs the solution as Eq. (23),

$$min_f[u(f, g) + \beta v(f)] \tag{23}$$

where f is sought to minimize $u(f, g) + \beta v(f)$. The first term, $u(f, g)$, describes how the real image data is related to the measured data. In other words, this term models the characteristic of the imaging system. The second term is the regularization term with the regularization operator v operating on the original image f, and the regularization parameter β used to tune up the weight of the regularization term. By adding the regularization term, the original ill-posed problem turns into a well-posed one, that is, the insertion of the regularization operator puts some constraints on what f might be, which makes the solution more stable.

In our problem, besides noise and blur, another corruption, missing data, is inserted to the measured image g. The consistency method using separable deblurring constructs a well-posed problem by putting several restrictions on the image data and the blur kernel. The MAP method is actually a Bayesian interpretation of regularization problems, which is solved by a global optimization technique, called mean field annealing (MFA).

7.8 Degradation and Distortion Characterization

Related parameters include degrees of radiometric and geometric distortions, variances of point spread function and noise distribution, and the amount of missing columns.

7.8.1 Degree of Radiometric Distortion

For vignetting-type radiometric distortion, the brightness degradation can be modeled by a series of concentric ellipses. We define the degree of radiometric distortion (η_r) by the eccentricity of the most inner ellipse (all the concentric ellipses have the same eccentricity). η_r is thus equal to ratio of the length of the major axis (a) to the length of the minor axis (b), as $\eta_r = a/b$.

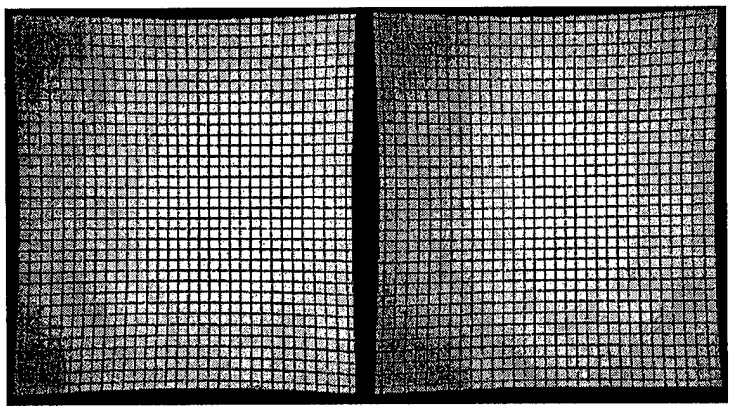


Figure 44. A real image used to compute the degree of both radiometric distortion and geometric distortion.

We use the image in Fig. 44 to compute η_r of both the left sensor image and the right sensor image as Eq. (24),

$$\eta_r(left) = \frac{777}{664} = 1.17$$
$$\eta_r(right) = \frac{705}{480} = 1.47 \tag{24}$$

7.8.2 Degree of Geometric Distortion

The degree of geometric distortion (η_g, mainly pincushion distortion) is measured at the smaller end of the fiber-optic taper. The η_g in the long direction is defined as Eq. (25) by the manufacturer, where L_c, L_m, L_{s1} and L_{s2} are illustrated in Fig. 45.

$$\eta_g = \frac{L_c - L_m}{L_{s1} + L_{s2}} \times 100 \tag{25}$$

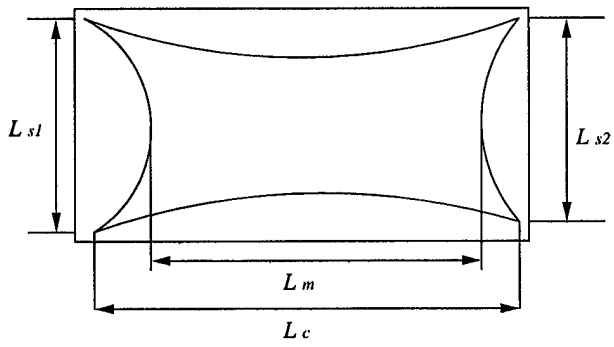


Figure 45. Parameter illustration in the definition of pincushion distortion degree.

By obtaining the corresponding L_{s1}, L_{s2}, L_m, and L_c from both the left and right parts of the original image in Fig. 44, we calculate the degree of pincushion distortion as Eq. (26).

$$\eta_g(left) = \frac{1075 - 1043}{1078 + 1175} \times 100 = 1.42\%$$
$$\eta_g(right) = \frac{1086 - 1049}{1171 + 1180} \times 100 = 1.57\% \tag{26}$$

7.8.3 Point Spread Function

The point spread function (PSF) can be estimated by the derivative of the edge spread function (ESF). Fig. 46 shows the ESFs and the corresponding PSFs along both horizontal and vertical directions at different locations (from both left and right parts of the image, select samples at top-left, top-middle, top-right, middle-left, middle-middle, middle-right, bottom-left, bottom-middle, and bottom-right respectively) of the image in Fig. 44, where the derivatives are taken numerically.

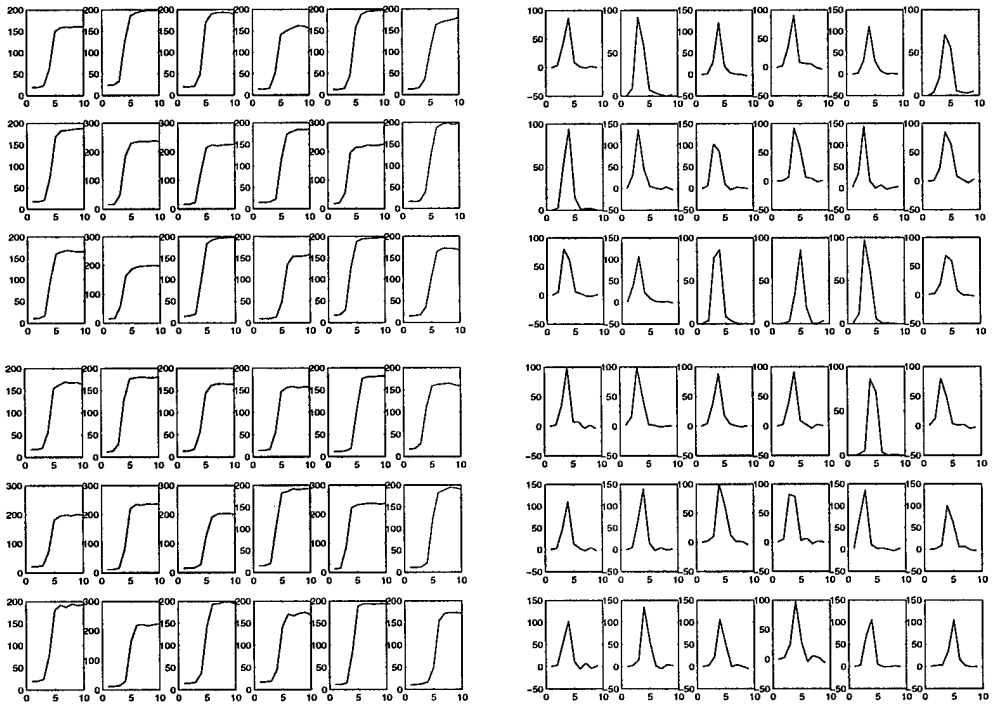


Figure 46. ESFs (left column) and PSFs (right column) along the horizontal direction (top) and the vertical direction (bottom).

We observe that the PSFs are all behave like a Gaussian, and the ESFs are like an error function. We therefore fit the sample data to the error function instead of the Gaussian.

The best fitting standard deviations (σ) for each of the error functions in Fig. 46 are listed in Table 4, and further illustrated in Fig. 47. We observe that the values of σ are very close at different positions and different edge orientations. In other words, the shapes of different ellipses in Fig. 47 are very similar, which means the PSF is roughly homogeneous. Therefore, we substitute $\sigma = 1.85$ to the 2-D Gaussian function (Eq. (27)), and sample it to create a 5 x 5 blur kernel (h') as Eq. (28). Kernel h' is further normalized to h as Eq. (29).

TABLE 4. The best fitting standard deviation for ESF.

		left part of the image			right part of the image		
along the horizontal direction		left	middle	right	left	middle	right
	top	1.82	1.85	1.86	1.85	1.85	1.94
	middle	1.83	1.84	1.83	1.83	1.84	1.84
	bottom	1.86	1.88	1.85	1.85	1.85	1.83
along the vertical direction	top	1.83	1.85	1.83	1.85	1.83	1.90
	middle	1.84	1.84	1.75	1.90	1.85	1.84
	bottom	1.84	1.83	1.85	1.82	1.83	1.82

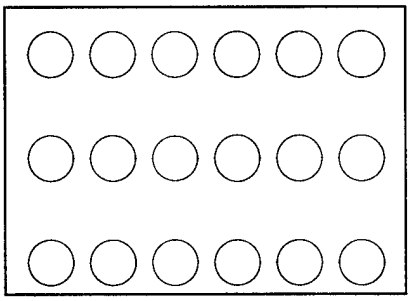


Figure 47. Illustration of the fitting σ (σ_x, σ_y) value from Table 4 by ellipses. The length of the axes of each ellipse is taken from σ along the horizontal direction, and σ along the vertical direction at the same location of the image. For example, the axes of the top-left ellipse is (1.82, 1.83) taken from cells (1, 1) and (3, 1) of the table.

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y^2}{2\sigma^2}\right) \tag{27}$$

$$h' = \begin{bmatrix} 0.0145 & 0.0224 & 0.0259 & 0.0224 & 0.0145 \\ 0.0224 & 0.0347 & 0.0402 & 0.0347 & 0.0224 \\ 0.0259 & 0.0402 & 0.0465 & 0.0402 & 0.0259 \\ 0.0224 & 0.0347 & 0.0402 & 0.0347 & 0.0224 \\ 0.0145 & 0.0224 & 0.0259 & 0.0224 & 0.0145 \end{bmatrix} \tag{28}$$

$$h = \begin{bmatrix} 0.0210 & 0.0326 & 0.0377 & 0.0326 & 0.0210 \\ 0.0326 & 0.0506 & 0.0585 & 0.0506 & 0.0326 \\ 0.0377 & 0.0585 & 0.0677 & 0.0585 & 0.0377 \\ 0.0326 & 0.0506 & 0.0585 & 0.0506 & 0.0326 \\ 0.0210 & 0.0326 & 0.0377 & 0.0326 & 0.0210 \end{bmatrix} \tag{29}$$

7.8.4 Noise Characterization

Noise is characterized by taking a homogeneous segment from the captured image, and observing its histogram distribution. Fig. 48 shows a flat frame captured from our imaging system. We select nine homogeneous segments at different locations (top-left, top-middle, top-right, middle-left, middle-middle, middle-right, bottom-left, bottom-middle, and bottom-right) from both the left and right parts of the flat frame and draw their histograms as Fig. 49. From Fig. 49, we observe that all the histograms behave like Gaussian. Again, we use *interopt* to fit the histograms to Gaussian and obtain the average best fitting standard deviation which is 3.9.

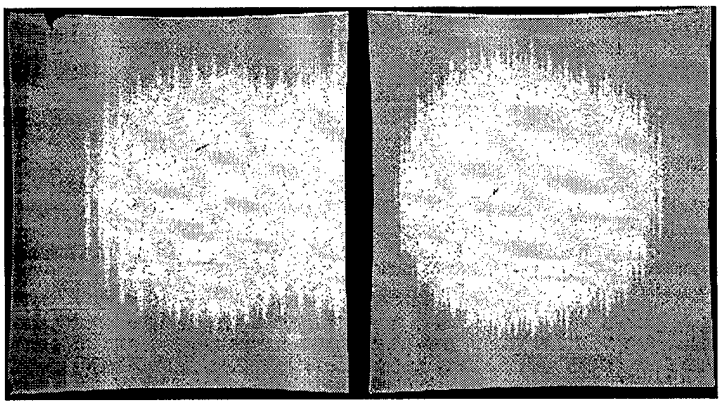


Figure 48. The image of a flat frame.

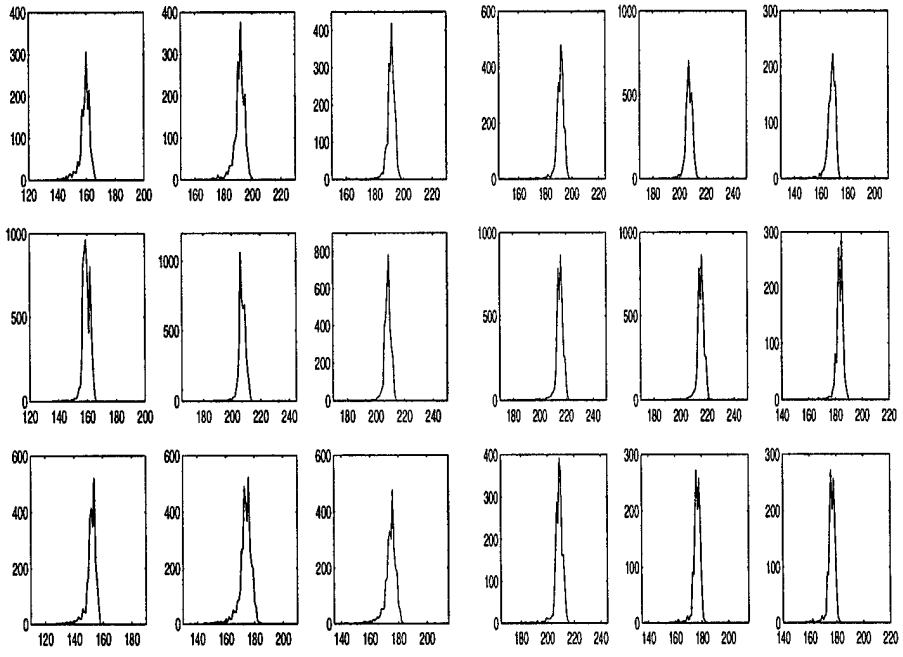


Figure 49. Noise distribution of a homogeneous segment from nine different locations of both the left (left column) and right (right column) parts of the image.

This method gives us a hypothetical idea about how noise contributes to the formation of the image.

7.8.5 Missing Data Identification

We measured that the thinnest line our CCD sensor can detect is about 50.8μm (0.002in), that is, the width of a pixel is 50.8μm. In order to identify how many columns are missing at the butting edge, we take images of a Precision Ronchi ruling across the butting edge and not across the butting edge.

Ronchi rulings are evenly spaced lines running parallel to each other with line width equal to space width. The ruling we use has 250 lines per inch. Therefore, the line width of the ruling can be calculated by Eq. (30). Since the width of each line in the ruling matches the width of the thinnest line the sensor can detect, the number of missing columns can be identified by counting the difference of number of lines from ruling image across the butting edge and ruling image fully contained in one sensor, as shown in Fig. 50.

$$\frac{1in}{250lines \times 2} \times 25.4cm = 0.0508cm = 50.8\mu m \tag{30}$$

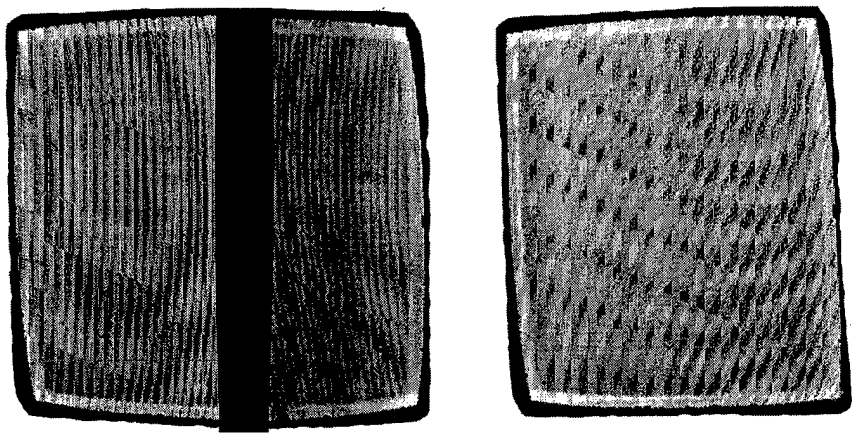


Figure 50. Images of Ronchi rulings across the butting edge and not across the butting edge.

We count number of lines across different positions of the ruling, with the results that the left image detects 237 lines, and the right image detects 238 lines. Therefore, we verify that the butted CCD sensor has one column missing at the butting edge.

7.9 The Consistency Method with Separable Deblurring

By putting some restrictions on input data, and the blur operator, we are able to transform the missing data estimation problem into a well-posed one. Several assumptions are made to simplify the problem. Possible relaxation of these assumptions are further discussed in Appendix 7.9.5.

- the blur kernel is separable;
- the blur kernel is exactly known;
- noise is not considered at this stage of study;
- only one column of data is missed in the original image; and
- the original image is of integer type.

The image formation is modeled as Eq. (31),

$$f \otimes h = g - n \tag{31}$$

where g is the measured image with one column of data missing, f is the original image, and n, the noise, is regarded as small perturbation of g and is neglected at this stage of study. The convolution kernel h is discrete and of finite support. Both f and g are M x N matrices.

The consistency method is proposed based on the assumption that the blur kernel is separable, such as Gaussian. The separability property of Gaussian leads to the separable deblurring where two N x N sparse matrices are generated based on the separated blur kernels. We demonstrate that this transformed problem is well-conditioned.

The consistency approach concerns three issues: separable deblurring, Householder triangularization based QR factorization (HHQR), and consistency in missing data estimation [Qi et al. 98].

7.9.1 Separable Deblurring

Since the convolution kernel h is discrete and of finite support, it can be written as a matrix like Eq. (32). If h is also separable, then the convolution can be performed separately as Eq. (33), where hy and hx are the separated vertical and horizontal components of h. $\bullet$ denotes the matrix multiplication, and $\otimes$ is the convolution operator.

$$h = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \cdots & \cdots & \cdots \\ h_{m1} & \cdots & h_{mn} \end{bmatrix} \tag{32}$$

$$f \otimes h = f \otimes (hy \bullet hx) = (f \otimes hy) \otimes hx = (f \otimes hx) \otimes hy \tag{33}$$

Assume f is an M x N image and h is an m x n Gaussian kernel, the convolution of image f with the horizontal kernel component hx can be equally achieved by a matrix multiplication between f and an N x N $\left(\left\lfloor \frac{n-1}{2} \right\rfloor, \left\lceil \frac{n-1}{2} \right\rceil\right)$-band matrix Dx (Eq. (34)), generated from hx; and the convolution with hy can be similarly achieved by a matrix multiplication with an M x M $\left(\left\lfloor \frac{m-1}{2} \right\rfloor, \left\lceil \frac{m-1}{2} \right\rceil\right)$-band matrix Dy (Eq. (35)), generated from hy.

$$Dx = \begin{cases} hx\left(\left\lfloor \frac{n+1}{2} \right\rfloor - i + j\right) & -\left\lceil \frac{n-1}{2} \right\rceil \le i - j \le \left\lfloor \frac{n-1}{2} \right\rfloor \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

$$Dy = \begin{cases} hy\left(\left\lfloor \frac{m+1}{2} \right\rfloor - i + j\right) & -\left\lceil \frac{m-1}{2} \right\rceil \le i - j \le \left\lfloor \frac{m-1}{2} \right\rfloor \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

For example, if we have an image f of dimension 5 x 4, and a separable blur kernel h of dimension 4 x 3:

$$f = \begin{bmatrix} f11 & f12 & f13 & f14 \\ f21 & f22 & f23 & f24 \\ f31 & f32 & f33 & f34 \\ f41 & f42 & f43 & f44 \\ f51 & f52 & f53 & f54 \end{bmatrix}, \; h = \begin{bmatrix} h11 & h12 & h13 \\ h21 & h22 & h23 \\ h31 & h32 & h33 \\ h41 & h42 & h43 \end{bmatrix} = \begin{bmatrix} hy1 \\ hy2 \\ hy3 \\ hy4 \end{bmatrix} \bullet \begin{bmatrix} hx1 & hx2 & hx3 \end{bmatrix},$$

the matrices Dx and Dy then look like:

$$Dx = \begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & hx3 & 0 \\ 0 & hx1 & hx2 & hx3 \\ 0 & 0 & hx1 & hx2 \end{bmatrix}, \quad Dy = \begin{bmatrix} hy2 & hy3 & hy4 & 0 & 0 \\ hy1 & hy2 & hy3 & hy4 & 0 \\ 0 & hy1 & hy2 & hy3 & hy4 \\ 0 & 0 & hy1 & hy2 & hy3 \\ 0 & 0 & 0 & hy1 & hy2 \end{bmatrix}.$$

With Dx and Dy, we can rewrite Eq. (33) into Eq. (36), which gives us another interpretation on how the original image is blurred. Eq. (37) is the inverse problem to Eq. (36), where if matrices Dx and Dy are both well-conditioned, the solution f should be stable. We will analyze the conditioning of Eq. (36) in detail in Appendix 7.9.4.

$$g = (f \otimes hx) \otimes hy = Dy \bullet (Dx \bullet f^T)^T \tag{36}$$

$$f = (Dx^{-1} \bullet (Dy^{-1} \bullet g)^T)^T \tag{37}$$

7.9.2 HHQR

To actually solve the system of Eq. (36), we use Householder triangularization based QR factorization (HHQR) algorithm [Trefethen and Bau 97]. HHQR has been proven to be *backward-stable* which means if the linear system is well-conditioned, this algorithm can provide the accurate solution. Assume the linear system is $Ax = b$, HHQR can then be implemented in the following three steps:

- using QR factorization to factor A into $Q \bullet R$, i.e., $QRx = b$ where Q is a unitary matrix and R is an upper triangular matrix generated by HH triangularization;
- since the inverse of a unitary matrix is equal to its transpose, $Rx = Q^{-1}b = Q^T b = y$;
- since R is an upper triangular matrix, we can solve x by simple scaler multiplication and subtraction without inverse operation.

In our problem, HHQR algorithm is used twice to solve the two linear systems derived from Eq. (36):

(1) solve $g = Dy \bullet f_y$ for f_y, and

(2) solve $f_y^T = Dx \bullet f^T$ for f.

7.9.3 Missing Data Estimation by Consistency

Assume the original image has only integer brightness values, and these values are within [0, 255], in particular, the missing pixels should have their original integer brightness values between 0 and 255, the two linear systems are solved as follows:

(1) Solve $g = Dy \bullet f_y$ for f_y using HHQR. The process is to deblur image g along the vertical direction. We show in Appendix 7.9.4 that matrix Dy is well-conditioned such that the inverse problem is well-posed.

(2) For the second linear system $f_y^T = Dx \bullet f^T$, HHQR can not be used directly, since there is one column of missing pixels in f_y. Instead, we solve each row of the image separately, i.e. solve $f_y^T(i) = Dx \bullet f^T(i)$, where i represents the ith row in that image. For the example we took at the end of Appendix 7.9.1, the second linear system can be written as Eq. (38),

$$\begin{bmatrix} f_y(i,1) \\ f_y(i,2) \\ f_y(i,3) \\ f_y(i,4) \end{bmatrix} = \begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & hx3 & 0 \\ 0 & hx1 & hx2 & hx3 \\ 0 & 0 & hx1 & hx2 \end{bmatrix} \bullet \begin{bmatrix} f(i,1) \\ f(i,2) \\ f(i,3) \\ f(i,4) \end{bmatrix} \tag{38}$$

Assume the missing pixel is in the 3rd column, we then have 5 unknowns ($f(i,1)$, $f(i,2)$, $f(i,3)$, $f(i,4)$ and $f_y(i,3)$) but only 4 equations in the linear system. To solve this problem, we need to find another condition. Since we already know that $f(i,3)$ should be an integer between 0 and 255, an exhaustive search can be carried on by assuming $f(i,3)$ is each one of them, which gives us 4 unknowns and 4 equations. Reconstruct vectors f_y, f, and matrix Dx in Eq. (38), so that all the unknowns are at one side of the equation, like Eq. (39). We denote the matrix reconstructed from Dx as Drx, and show in Appendix 7.9.4 that it is also well-conditioned.

$$\begin{bmatrix} f_y(i,1) \\ f_y(i,2) - hx3 \cdot f(i,3) \\ -hx2 \cdot f(i,3) \\ f_y(i,4) - hx1 \cdot f(i,3) \end{bmatrix} = \begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & 0 & 0 \\ 0 & hx1 & -1 & hx3 \\ 0 & 0 & 0 & hx2 \end{bmatrix} \bullet \begin{bmatrix} f(i,1) \\ f(i,2) \\ f_y(i,3) \\ f(i,4) \end{bmatrix} \tag{39}$$

Solve the new linear system by HHQR, and check the solution to see if it satisfies the *consistency criterion*, that is, the solutions ($f(i,1)$, $f(i,2)$, $f(i,4)$) are all between 0 and 255 and are all integers. If they do, then we can go on and solve the next row; otherwise, next value of $f(i,3)$ is attempted. The well-conditioned problem and the backward-stable algorithm assure that the solution is unique and accurate.

How to choose the initial value of the missing pixel plays an important role in speeding up the exhaustive search. We use the average brightness of the missing pixel's left and right neighbors as the first attempt, and pick up the next attempt by swinging around this value with increasing step (Fig. 51). In practice, this saves half of the time compared to the normal approach which always starts from brightness 0.

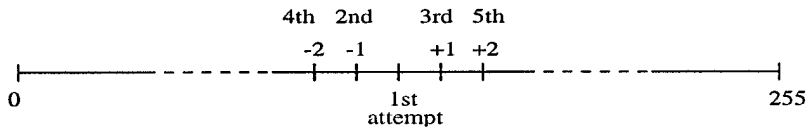


Figure 51. Choose initial value for the missing pixel in exhaustive searching.

7.9.4 Conditioning Analysis

Conditioning of a mathematical problem is measured by the sensitivity of output to changes in input. For a well-conditioned problem, a small change of input does not affect the output much; while for an ill-conditioned problem, a small change of input can change the output a great deal.

Condition number is the measurement of the conditioning of a problem. Generally, it is defined as Eq. (40). The larger the condition number, the more ill-conditioned the problem is.

$$\text{condition number} \approx \frac{\text{change in output}}{\text{change in input}} \tag{40}$$

The conditioning of a linear system $Ax = b$ is determined by the condition number of matrix A. The relative condition number K is defined as Eq. (41),

$$K = \|A\|\|A^{-1}\| \tag{41}$$

where $\|.\|$ usually indicates the 2-norm. K is in the range of $[1, \infty)$. When $K \gg 1$, the linear system is ill-conditioned.

While most of the literature on image restoration concentrates on smoothing methods to reduce the effect of noise, few [Forbes and Anh 94][Milinazzo et al. 87] pay attention to quantify the amount of ill condition of the blur kernel. Since the whole idea of the consistency method is built upon the well-conditioning of the two linear systems, analyzing the amount of ill condition of matrices becomes an important issue.

In the missing data estimation problem, the conditioning of the two linear systems is measured by the condition number of matrices *Dx*, *Dy*, and *Drx*. Since *Dx* and *Dy* has similar structure, we only analyze *Dy* and *Drx*.

Condition number of *Dy*. We compute the condition number of matrix *Dy* generated with different image sizes and Gaussian blur kernel sizes. The results are listed in Table 5. We can see that for kernel size larger than 3, the condition number does not change much when increasing the image size. Also, when increasing the kernel size, the condition number always fluctuates around 10.

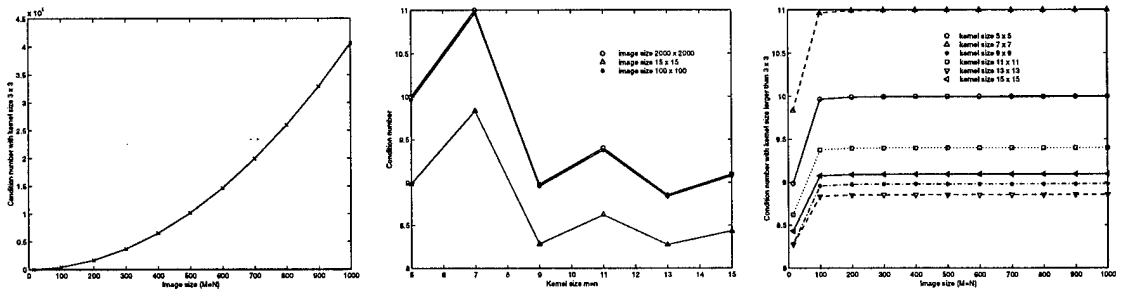
TABLE 5. Condition number of matrices *Dy*.

Image Size (*N* x *N*)	Condition Number of Different Kernel Sizes (*n* x *n*)					
	3	5	7	9	11	13
15	48.3742	9.9852	9.8348	8.2826	8.6213	8.2726
100	4133.6000	9.9659	10.9641	8.9596	9.3753	8.4998
1000	406100.0000	9.9996	10.9996	8.9795	9.3997	8.8548
2000		9.9999	10.9999	8.9797	9.3999	8.8550

Fig. 52 is several plots illustrating the rate of growth of the condition number with respect to different image sizes and kernel sizes. Fig. 52 (a) is the growth curve for 3 x 3 blur kernel. The curve increases steadily without converging to an upper bound, which indicates the system is ill-conditioned.

Fig. 52 (c) shows six curves with respect to six different kernel sizes: 5 x 5, 7 x 7, 9 x 9, 11 x 11, 13 x 13, and 15 x 15. All the curves behave similar: a sharp increase at the beginning and asymptotically converge to a constant (less than 11) after the image size is larger than 100 x 100. Fig. 52 (b) displays three curves corresponding to different image sizes. It indicates that all the condition numbers fall into the area with an upper bound created by the largest image size (2000 x 2000) and a lower bound related to the smallest image size (15 x 15). The upper bound is around 10, which demonstrates the well-conditioning of the system.

From the above analysis, we claim that when blur kernel is larger than 3 x 3, the first linear system $g = Dy \bullet f_y$ is well-conditioned.

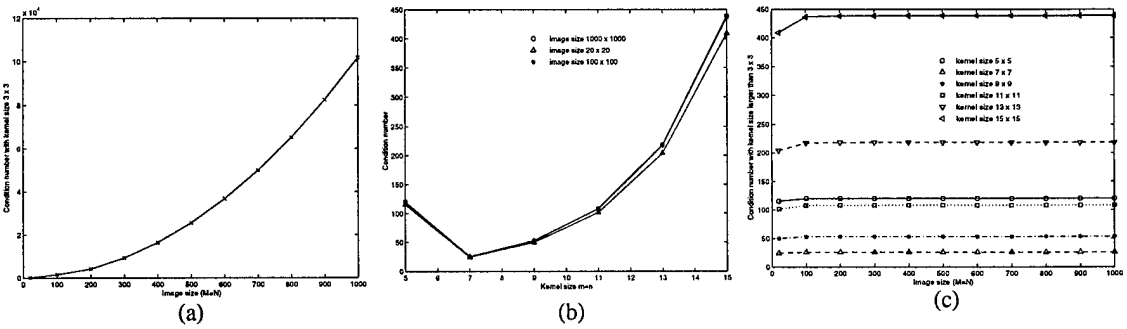


(a) (b) (c)

Figure 52. Condition number of matrix *Dy* with respect to different image and kernel sizes.

Condition number of *Drx*. The condition number of *Drx* indicates the conditioning of the second linear system (Eq. (39)). *Drx* is transformed from *Dx*, with only one column different. If the missing column in the original image is column *j*, then *Drx* is constructed by setting the *j*th column of *Dx* to zero except element (*j*, *j*) with a value -1, as indicated in Eq. (42), where $j = 3$.

$$\begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & hx3 & 0 \\ 0 & hx1 & hx2 & hx3 \\ 0 & 0 & hx1 & hx2 \end{bmatrix} \rightarrow \begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & 0 & 0 \\ 0 & hx1 & -1 & 0 \\ 0 & 0 & 0 & hx2 \end{bmatrix} \tag{42}$$



(a) (b) (c)

Figure 53. Condition number of matrix *Drx* with respect to different image and kernel sizes.

Fig. 53 is three plots similar to Fig. 52. For kernel size larger than 3 x 3, the condition number still converges asymptotically to a constant after an apparent increase at the beginning. Although the constant is larger (less than 450) than that from Fig. 52, it still indicates well-conditioning. Fig. 53 (b, c) also tells us that the condition number is more affected by different kernel size than different image size, which can be seen from the close distance between the three curves in Fig. 53 (b) and the comparatively larger distance between curves in Fig. 53 (c).

7.9.5 Assumption Relaxation Analysis

At the beginning of Appendix 7.9, we made several assumptions to simplify the restoration problem. This section discusses the possible relaxations to these assumptions. Except that the blur kernel still needs to be separable, we show that other assumptions can all be relaxed to a certain degree, though by sacrificing certain amount of accuracy of the solution. We first analyze sensitivity of solution to perturbations in blur kernel (h), and perturbations in measured image (g) caused by the insertion of noise (n). We also evaluate the conditioning of problem when there are more than one column of missing data. Finally, we relax the assumption of integer-typed original image, and design a *neighbor least square error* criterion to select the optimal solution.

Sensitivity of Solution to Perturbations in Blur Kernel. According to [Trefethen and Bau 97], in a linear system $Ax = b$, sensitivity of solution x to perturbations in b can be measured by Eq. (43),

$$\frac{\|x - \tilde{x}\|}{\|x\|} = \frac{K(A)}{\eta}\frac{\|\delta b\|}{\|b\|} \tag{43}$$

where $\eta = \frac{\|A\| \cdot \|x\|}{\|Ax\|}$. Sensitivity of x to perturbations in A can be estimated by Eq. (44).

$$\frac{\|x - \tilde{x}\|}{\|x\|} \le K(A)\frac{\|\delta A\|}{\|A\|} \tag{44}$$

In our system, perturbations in blur kernel (h) are reflected in blur matrices *Dy* and *Drx* as shown in Eq. (45) and Eq. (46).

$$(Dy + \delta Dy)\tilde{f}_y = g \tag{45}$$

$$(Drx + \delta Drx)\tilde{f}^T = f_y^T \tag{46}$$

Based on Eq. (43) and Eq. (44), sensitivity of f with respect to perturbations in Dx can be expressed by Eq. (47), and sensitivity of f to perturbations in Dy by Eq. (48),

$$\frac{\|f-\tilde{f}\|}{\|f\|} \le K(Drx)\cdot\frac{\|\delta Drx\|}{\|Drx\|} \tag{47}$$

$$\frac{\|f-\tilde{f}\|}{\|f\|} = \frac{K(Drx)}{\eta_x}\cdot\frac{\|\delta f_y\|}{\|f_y\|} \le \frac{K(Drx)K(Dy)}{\eta_x}\cdot\frac{\|\delta Dy\|}{\|Dy\|} \tag{48}$$

where $\eta_x = \dfrac{\|Drx\|\cdot\|f\|}{\|Drx\cdot f\|}$, a number greater than 1, but very close to 1. If Drx and Dy have same degree of perturbation, then Eq. (47) derives a smaller upper bound than Eq. (48). Therefore, sensitivity of f to perturbations in blur kernel h is measured more accurate by Eq. (47).

If $K(Drx)\approx 10^{c}$, and $\|\delta Drx\|/\|Drx\|\approx 10^{-d}$, when $c = d$, sensitivity of solution is close to 1 (Eq. (49)), which means the problem is ill-conditioned, since we expect no correct digits from the output f.

$$\frac{\|f-\tilde{f}\|}{\|f\|} = \frac{\|\delta f\|}{\|f\|} \le 1 \tag{49}$$

Sensitivity of Solution to Perturbations in Measured Image. Perturbations in measured image g can be interpreted as the effect of inserted noise. The two perturbed linear systems can be written as Eq. (50) and Eq. (51).

$$Dy\cdot\tilde{f}_y = g+\delta g \tag{50}$$

$$Drx\cdot\tilde{f} = f_y+\delta f_y \tag{51}$$

Sensitivity of f_y to perturbations in measured image g is defined by Eq. (52) based on Eq. (43),

$$\frac{\|f_y-\tilde{f}_y\|}{\|f_y\|} = \frac{K(Dy)}{\eta_y}\cdot\frac{\|\delta g\|}{\|g\|} \tag{52}$$

where $\eta_y = \dfrac{\|Dy\|\cdot\|f_y\|}{\|Dy\cdot f_y\|}$, a number greater than 1, but very close to 1. Sensitivity of f to perturbations in measured image g can be derived by Eq. (53).

$$\frac{\|f-\tilde{f}\|}{\|f\|} = \frac{K(Drx)}{\eta_x}\cdot\frac{\|\delta f_y\|}{\|f_y\|} = \frac{K(Drx)K(Dy)}{\eta_x\cdot\eta_y}\cdot\frac{\|\delta g\|}{\|g\|} \tag{53}$$

Compare Eq. (53) with Eq. (47), if Drx and g have same degree of perturbation, then solution f can be less sensitive to perturbations in blur kernel than to perturbations in measured image.

Conditioning Analysis for More Than One Missing Column. If there are more than one column of missing data in the measured image, condition number of the vertical blur matrix Dy stays the same, but condition number of the reconstructed horizontal blur matrix Drx is changed. Eq. (54) shows a template of Drx transformed from Dx when two columns of data (the 2nd and the 3rd columns) are missed. The missed ith and jth columns of Dx are set to zeros except at element (i, i) and (j, j) with a value -1, similar to the formation in Eq. (42).

$$\begin{bmatrix} hx2 & hx3 & 0 & 0 \\ hx1 & hx2 & hx3 & 0 \\ 0 & hx1 & hx2 & hx3 \\ 0 & 0 & hx1 & hx2 \end{bmatrix} \rightarrow \begin{bmatrix} hx2 & 0 & 0 & 0 \\ hx1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & hx2 \end{bmatrix} \tag{54}$$

With more than one column missing, two problems need to be considered: how many missing columns can be in the image, and how far away these missing columns should be apart for stable inverse. We still use condition number to measure the conditioning of the problem.

Fig. 54 shows the condition number of *Drx* with blur kernel (*h*) at dimension 5 x 5. The three plots are generated with different numbers of missing columns (2, 3, and 4). We can see that no matter how many columns are missed in the measured image, and how far away the two nearest missing columns are apart, all of the plots behave similarly - after an initial perturbation, all plots converge to a constant condition number, which is around 120.

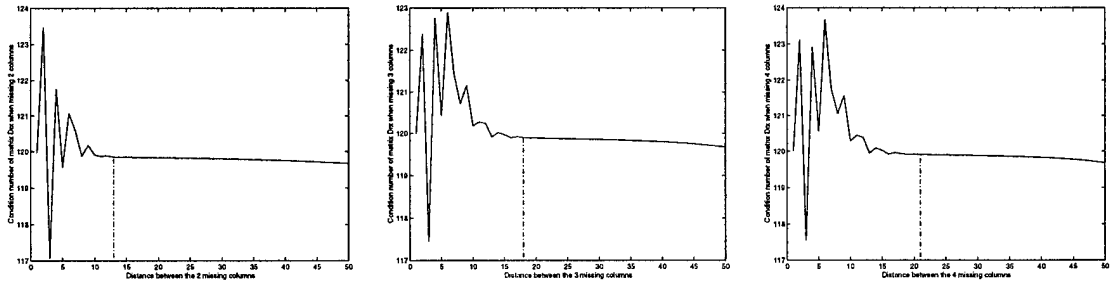


Figure 54. Condition number of *Drx* with different numbers of missing columns at kernel size 5 x 5.

Fig. 55 shows the condition number of *Drx* with blur kernel (*h*) at dimension 7 x 7. The three plots also have similar characteristics as those from Fig. 54. The difference lies in that after an initial perturbation, the three plots in Fig. 55 converge to a larger constant condition number, which is around 260.

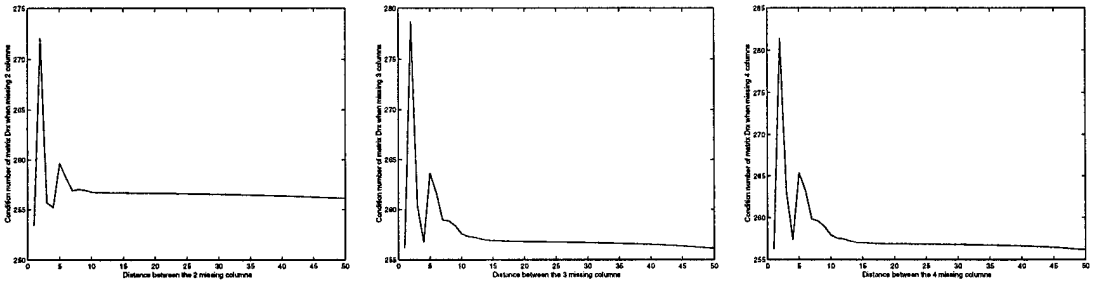


Figure 55. Condition number of *Drx* with different numbers of missing columns at kernel size 7 x 7.

Fig. 56 compares the condition number of *Drx* with different kernel size when two columns of data are missed. Both of the two curves increase steadily with respect to the kernel size.

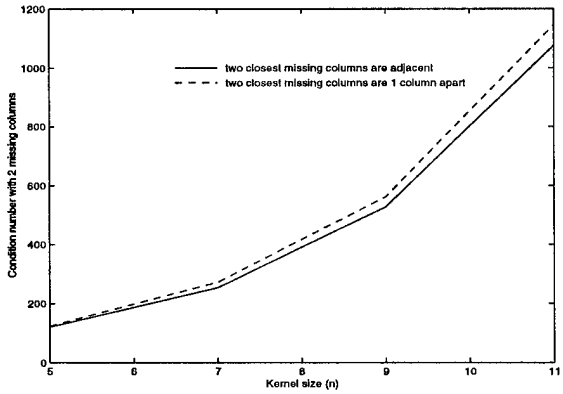


Figure 56. Condition number comparison of *Drx*, missing 2 columns of data, with different kernel size: 5 x 5, 7 x 7, 9 x 9, and 11 x 11.

These results substantiate our claim that when more than one column of data are missed, it is not the image size, or the distance between two nearest missing columns, or the number of missing columns that affect the condition number, it is the *kernel size* that plays the most important role. When the kernel size is less than 10 x 10, the problem is still well-conditioned. The larger the kernel size, the worse the problem is conditioned.

Neighbor Least Square Error Consistency Criterion. The last assumption we made is that the original image needs to be of integer type. We relax this assumption to float-type image and design a new consistency criterion to be used when selecting the optimal solution at the last step of the consistency method. We call it *neighbor least square error consistency* criterion. It works in the following steps:

1) Estimate the missing column in the measured image using average value of its left and right neighbors;

2) Apply HHQR to solve the first linear system $g = Dy \bullet f_y$ for f_y;

3) Reconstruct horizontal blur matrix *Dx* to *Drx*, and solve the second linear system of Eq. (39) one row at a time.

4) For each possible solution of $f(i)$ (the ith row of the restored image), compute the neighbor least square error by Eq. (55), where d is the column position of the missing data, and δ is the number of neighbors that is involved in this least square error computation. It is important to choose close neighbors instead of the entire row for the computation.

$$\left\| f(i) \otimes hx - f_y(i) \right\|^2 = \sum_{j=d-\delta}^{d+\delta} \left[(f(i) \otimes hx)_j - f_y(i,j) \right]^2 \tag{55}$$

This criterion relaxed the integer-type assumption, unfortunately, the search in the fourth step has to be exhaustive. That is, we need to assume the value of the missing pixel to be from 0 to 255. Based on the 256 solution, we select the one that minimizes Eq. (55).

7.10 MAP Using MFA

This is another approach we proposed to solve the problem of missing data estimation with denoising and deblurring. This approach uses the optimization technique, MFA, to solve an MAP estimate. With the image formulated as Eq. (24), the maximum *a-posteriori* probability (MAP) approach tries to find an estimate of image f that maximizes the *a-posteriori* probability $p(f|g)$ as Eq. (56).

$$\hat{f} = \arg max_f p(f|g) \tag{56}$$

According to Bayes' rule, $p(f|g)$ can be written as Eq. (57),

$$p(f|g) = \frac{p(g|f)P(f)}{P(g)} \tag{57}$$

where $P(f)$ is the *a-priori* probability of the unknown image f, $P(g)$ is the probability of g which is a constant when g is given, and $p(g|f)$ is the conditional probability density function (pdf) of the observation g. In [Szeliski 89], $P(f)$ is called the *prior model*, and $p(g|f)$, the *sensor model*, which is a description of the noisy or stochastic processes that relate the original unknown image f to the measured image g.

Based on Bayes' rule, the MAP approach can then be derived as Eq. (58), where we neglect the constant $P(g)$.

$$\hat{f} = \arg max_f p(f|g) = \arg max_f (p(g|f)P(f)) \tag{58}$$

The MAP approach may be regarded as a Bayes interpretation to the regularization theory [Demoment 89][Li 95]. It eventually formulates an image restoration problem into finding the optimal solution of an

objective function which is very similar to the traditional regularization method, but derived from a different point of view.

In the next four subsections, we will discuss our design of the sensor model and the prior model. The mean field annealing global optimization technique is adopted here to obtain the restored image f which maximizes the *a-posteriori* probability.

7.10.1 Sensor Model

If we define the δ-function $\delta(i)$ as Eq. (59), then the sensor model can be expressed as Eq. (60),

$$\delta(i) = \begin{cases} 1 & i \in \Omega - \Omega_d \\ 0 & i \in \Omega_d \end{cases} \tag{59}$$

$$n = \{n_i | n_i = \delta(i) \cdot [(f \otimes h)_i - g_i]\} \tag{60}$$

where i denotes individual pixel, d denotes locations of missing data, Ω is the set of pixels of the entire image, and Ω_d is the set of missing pixels.

Assume noise is independent Gaussian noise with zero mean, i.e., $n_i \sim N(0, \sigma^2)$, then the likelihood density function can be written as Eq. (61), a product over the pixels of the image.

$$p(g|f) = \prod_{i \in \Omega - \Omega_d} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{[(f \otimes h)_i - g_i]^2}{2\sigma^2}\right) \cdot \prod_{i \in \Omega_d} \frac{1}{\sqrt{2\pi}\sigma} \tag{61}$$

7.10.2 Prior Model

The prior model is a probability description of the original image. It plays the same role as the regularization term: by providing some prior knowledge of the original image, so to put some constraints in the solution, the original *ill-posed* image restoration problem turns into a *well-posed* one.

The prior knowledge of the original image refers to the *a-priori* belief that the state of a pixel is entirely determined by the states of its neighboring pixels. Specifically, it is expected that pixels close to each other tend to have the same or similar brightness values [Besag 74][Besag 86]. [Geman and Geman 84] uses Markov Random Field (MRF) to represent the local property of the image.

An MRF is a probabilistic process in which all interaction is local. It is defined over a discrete field where the probability of a particular variable f_i defends only on a small number of its neighbors, which is expressed as Eq. (62).

$$p(f_i|f) = p(f_i|\{f_j, j \in \aleph_i\}) \tag{62}$$

Although an MRF is the correct model to represent the local property in the image, it is difficult to estimate either the conditional Markov distribution $p(f_i|f)$, or the joint probability directly from the conditional distribution. Fortunately, [Hammersley and Handscomb 64] and later paper [Besag 74] have shown the equivalence between Gibbs distributions and MRF. This equivalence allows the modeling of local structure through energies which describe the interactions of pixels within each clique of the neighborhood. The *a-priori* probability of an image by a Gibbs distribution is defined as Eq. (63),

$$P(f) = \frac{\exp(-U(f)/T)}{Z} \tag{63}$$

where $Z = \sum_f \exp(-U(f)/T)$ is a normalizing constant, called the *partition function*; T is the temperature of the model; and $U(f)$ is the *energy function*, that can be written as Eq. (64),

$$U(f) = \sum_i V_{C_i}(f) \tag{64}$$

where C_i is the clique formed involving pixel i. $V_{C_i}(f)$ is called the *potential*, which depends on the local configuration of clique C_i.

The prior energy function is usually formulated based on the smoothness property of the original image. Since the more probable configurations are those with higher $P(f)$ (thus lower $U(f)$), $U(f)$ should measure the extent to which the smoothness is violated [Li 95]. For spatially continuous MRFs, the energy function often involves derivatives. Different orders of derivatives imply different classes of smoothness. Based on the discussions in [Li 95], we summarize the different formulations of the energy function according to different kinds of image surfaces in Table 6:

TABLE 6. Prior energy for different surfaces.

Surface property	Derivative	Prior energy $U(f)$	
flat surface: $f(x, y) = a_0$	zero 1st-order derivative: $f_x = 0, f_y = 0$	$\iint (f_x^2 + f_y^2)dxdy$	membrane
planar surface: $f(x, y) = a_0 + a_1 x + a_2 y$	zero 2nd-order derivative: $f_{xx} = 0, f_{yy} = 0$ $f_{xy} = 0$	quadratic variation: $\iint (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)dxdy$	plate
		square Laplacian: $\iint (f_{xx} + f_{yy})^2 dxdy$	
quadratic surface: $f(x, y) = a_0 + a_1 x^2 + a_2 xy + a_3 y^2$	zero 3rd-order derivative: $f_{xxx} = 0, f_{yyy} = 0$ $f_{xxy} = 0, f_{xyy} = 0$	$\iint (f_{xxx}^2 + 3f_{xxy}^2 + 3f_{xyy}^2 + f_{yyy}^2)dxdy$	
		$\iint (f_{xxx}^2 + f_{yyy}^2)dxdy$	

[Blake and Zisserman 87] designed a clipped parabola (Fig. 57) to model the energy of interaction between neighbors in the weak string (the 1-D case of membrane), where ∇ represents any operator which returns a measure of the local "edginess" of the image, such as the square Laplacian. The central dip of this energy function punishes the difference $\nabla_i(f)$, and the plateaus allow discontinuity. Since both noise and edges can increase the brightness difference between neighbor pixels, and the model should only punish the noise but not the edges, the energy curve turns into a constant after $\nabla_i(f)$ surpasses a certain threshold. This model is in general not convex, therefore, graduated nonconvexity (GNU) algorithm is developed which approximates the energy with a piecewise smooth function as Eq. (65),

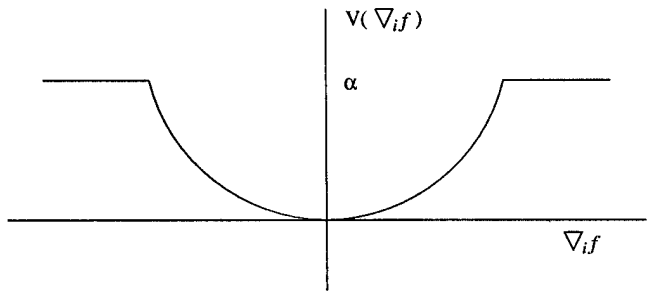


Figure 57. Prior energy of the GNC algorithm.

$$\tilde{V}(t) = \begin{cases} \lambda^2 t^2 & \text{if } (|t| < q) \\ \alpha - c'(|t| - r)^2/2 & \text{if } (q \le |t| < r) \\ \alpha & \text{if } (|t| \ge r) \end{cases} \tag{65}$$

where $r^2 = \alpha\left(\frac{2}{c'} + \frac{1}{\lambda^2}\right)$, and $q = \frac{\alpha}{\lambda^2 r}$.

In our system, we adopt the energy model proposed in [Bilbro and Snyder 88a], where the penalty function of Fig. 57 is interpreted as a Gaussian with the upside down, shown in Fig. 58. The prior energy function can then be written as Eq. (66),

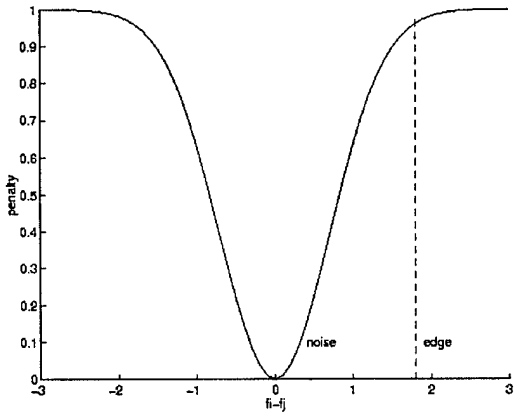


Figure 58. An ideal curve to represent energy function for the prior model (the horizontal axis is the difference in brightness of adjacent pixels).

$$U(f) = \sum_i -\frac{\beta}{\sqrt{2\pi}\tau} \exp\left(-\frac{(\nabla^k f_i)^2}{2\tau^2}\right) \tag{66}$$

where β is the parameter used to adjust how smooth the image goes, $\nabla^k f_i$ is some norm of the k-th derivative, and depends upon the property of the image. The derivative operation can also be formulated into a convolution as $(f \otimes r)^2$, with appropriate convolution kernel r. Therefore, the *a-priori* probability can be written as Eq. (67).

$$P(f) = Z^{-1} \exp\left(-\frac{\sum_i -\frac{\beta}{\sqrt{2\pi}\tau} \exp\left(-\frac{(f \otimes r)_i^2}{2\tau^2}\right)}{T}\right) \tag{67}$$

7.10.3 Objective Function

With both $p(g|f)$ and $P(f)$ developed, we can derive our objective function as Eq. (68) from the MAP estimate by taking the natural logarithms of Eq. (58); and by changing the sign, we convert the problem from maximizing a probability to minimizing an objective function.

$$H = -\ln(p(g|f)P(f)) = -\ln(p(g|f) - \ln P(f)) \equiv H_n + H_p \tag{68}$$

where

$$H_n = \frac{1}{2\sigma^2} \sum_{i \in \Omega - \Omega_d} [(f \otimes h)_i - g_i]^2 \tag{69}$$

and

$$H_p = -\sum_i \frac{\beta}{\sqrt{2\pi}\tau} \exp\left(-\frac{(f \otimes r)_i^2}{2\tau^2}\right) \tag{70}$$

H_n is called the *noise term* (Eq. (69)) which concerns both the measured image g and the original image f. H_p is the *prior term* (Eq. (70)), which is determined only by f. A global optimization technique called *mean field annealing* is used to find f which minimizes the objective function H globally.

7.10.4 Optimization by Mean Field Annealing

Mean field annealing (MFA) is used to find image f which minimizes the objective function H. MFA was introduced in late 80s by different groups working independently [Bilbro and Snyder 88a][Bilbro et al. 89][Peterson and Soderberg 89]. It is derived based on the principles of simulated annealing (SA). SA is able to locate the global minimum/maximum instead of being trapped in local minima/maxima because it allows uphill moves with a certain probability. Both MFA and SA relate to the physical process of annealing, where at high temperatures, the particles in an object are more likely to randomly change their states; when temperature is gradually reduced, particles remain comparatively stable, and a minimum energy state is achieved at last. The difference between MFA and SA is that SA stochastically simulates this process by random search, while MFA analytically approximates this process with a series of deterministic gradient descents [Bilbro et al. 92]. Therefore, MFA is much more efficient than SA, as much as 50 times faster [Bilbro et al. 89].

Since its introduction, MFA has found its applications in restoration of locally-homogeneous images [Hiriyannaiah et al. 89], range images [Bilbro and Snyder 89], and locally smooth images [Bilbro and Snyder 90]; in image segmentation [Snyder et al. 91]; in motion analysis [Abdelqader et al. 92]; in sensor fusion [Bilbro and Snyder 88b]; in image resolution increasing [Wang 96], and in solving anisotropic diffusion problems [Qi et al. 97].

In order to use MFA in our problem, the derivative of H with respect to f in the gradient descent equation (Eq. (71)) need to be derived. Eq. (72) shows the derivation.

$$f^{k+1} = f^k - \alpha \frac{\partial H}{\partial f} \tag{71}$$

$$\begin{aligned} \frac{\partial H}{\partial f_i} &= \frac{\partial H_n}{\partial f_i} + \frac{\partial H_p}{\partial f_i} \\ &= \frac{(n \otimes h_{rev})_i}{\sigma^2} + \left\{ \left(\frac{\beta (f \otimes r)}{\sqrt{2\pi}\tau^3} \exp\left(-\frac{(f \otimes r)^2}{2\tau^2}\right) \right) \otimes r_{rev} \right\}_i \end{aligned} \tag{72}$$

where h_{rev} and r_{rev} are the reversed kernel of h and r. n is defined in Eq. (60) where n_i is zero for unmeasured elements. The reverse convolution with h_{rev} will be discussed in detail in Appendix 7.10.5.

MFA is implemented by replacing parameter τ in the Gibbs distribution with $\tau + T$, where T is called the *temperature*. Initially, T is very large which results in a convex function minimized by $f_i = g_i$. Gradient descent uses this point as the starting point. By gradually reducing T, and making it approach zero, a minimum energy state will finally be achieved which is where the optimal solution locates.

7.10.5 Missing Data Estimation by MFA

Compare to the consistency method, MFA is more flexible, less vulnerable to large amount of noise and the inaccuracy calculation of blur kernel. Three issues need to be discussed in order to implement this approach:

(1) the choice of the blur kernel (h); (2) the choice of the neighborhood operation kernel (r); and (3) the reverse convolution with h_{rev} in the noise term.

The blur kernel (h). How to compute the blur kernel in an optimal way is beyond the scope of this research. [Hussain 97] states in detail about how to estimate the point spread function with subpixel accuracy. Here, we use the edge spread function (ESF) to estimate the point spread function (PSF). We first choose several sets of sample pixels around a horizontal edge and a vertical edge respectively from the measured image. Averaging the brightness of the sets of sample pixels to obtain a discrete data set of ESF in the horizontal and vertical direction, denoted as $\left\{ f^h_{i,j} \middle| j \in [e^v - \delta, e^v + \delta] \right\}$, and $\left\{ f^v_{i,j} \middle| i \in [e^h - \delta, e^h + \delta] \right\}$, where e^v and e^h are the locations of the vertical and horizontal edges from the measured image, δ is a small distance away from the edge location. The data sets $\{f^h_{i,j}\}$ and $\{f^v_{i,j}\}$ should behave like Fig. 59 (a), shape of an error function. The derivatives (or difference in discrete case) of $\{f^h_{i,j}\}$ and $\{f^v_{i,j}\}$ (denoted as $\{df^h_{i,j}\}$ and $\{df^v_{i,j}\}$) should behave like Fig. 59 (b), shape of a Gaussian.

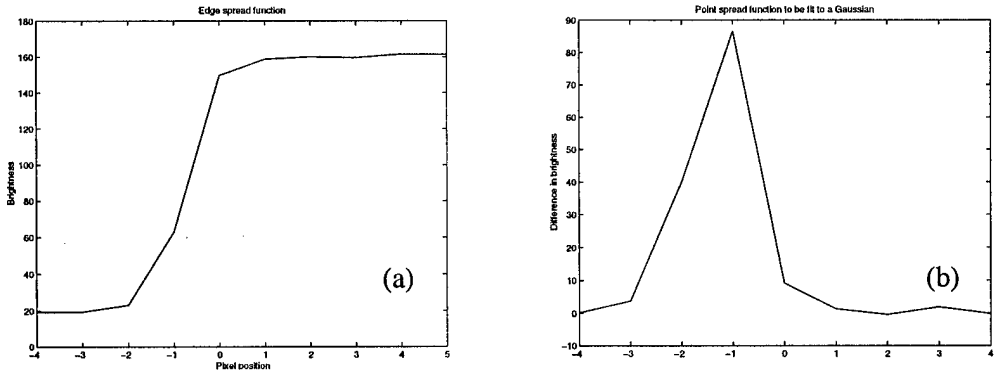


Figure 59. The edge spread function and the point spread function.

The PSF is characterized by fitting the discrete data sets $\{df^h_{i,j}\}$ and $\{df^v_{i,j}\}$ to a Gaussian, thus to obtain the standard deviations. Based on the relationship between an error function and a Gaussian function (expressed in Eq. (73)), we can instead fit the error function directly with the sampled data sets $\{f^h_{i,j}\}$ and $\{f^v_{i,j}\}$. By doing so, we avoid the difference operation and achieve higher accuracy in parameter estimation.

$$G(x) = \frac{a}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(t-\mu)^2/(2\sigma^2)} dt = \frac{a}{2}\left(1 + erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right) \tag{73}$$

With the standard deviations solved, we can then obtain the 2-D PSF, which is shown in Eq. (74). The continuous PSF is sampled to construct a 5 x 5 blur kernel h. Here, we use a software package called *interopt* [Bilbro and Snyder 91] to fit the sample data to an error function.

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{\sigma_y^2 x^2 + \sigma_x^2 y^2}{2\sigma_x^2\sigma_y^2} \right) \tag{74}$$

The neighborhood operation kernel (r). We choose quadratic variation (QV) to model the prior energy function. Compared to squared Laplacian, QV does not go to zero when $\frac{\partial^2 f}{\partial x^2} = -\frac{\partial^2 f}{\partial y^2}$ (a saddle point) and

conclude that a saddle point is a plane as Laplacian does. QV has a more stable description about the surface, it also simulates the second derivative operation, and permits piecewise linear solutions.

The discrete form of quadratic variation can be expressed as Eq. (75),

$$\sum_{i,j}[f_{xx}^2(i,j) + 2f_{xy}^2(i,j) + f_{yy}^2(i,j)] \tag{75}$$

where $f_{xx} = f_{i,j-1} + f_{i,j+1} - 2f_{i,j}$, $f_{xy} = -f_{i,j} - f_{i+1,j+1} + f_{i,j+1} + f_{i-1,j}$, and $f_{yy} = f_{i-1,j} + f_{i+1,j} - 2f_{i,j}$ [Blake and Zisserman 87]. All of the three 2nd derivatives can be constructed into a convolution operation with certain kernels h_{xx}, h_{xy}, and h_{yy} as Eq. (76).

$$h_{xx} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, h_{xy} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}, h_{yy} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \tag{76}$$

The reverse convolution in noise term. In Eq. (72), the convolution with reverse kernel (h_{rev}) is different from others because g has certain amounts of data missed. The convolution should be taken at places where a pixel has a measurement. Snyder derived the formulation for reverse convolution in [Snyder 99], which is briefly described as follows:

Take the 1-D example, assume the measured image has every other column of data missing as indicated in Fig. 60, where f_i is a pixel from the original image, g_i is a pixel from the measured image, and h is the horizontal blur kernel with a finite kernel size 5. We explain the derivative of the noise term (Eq. (72)) in gradient descent in detail.

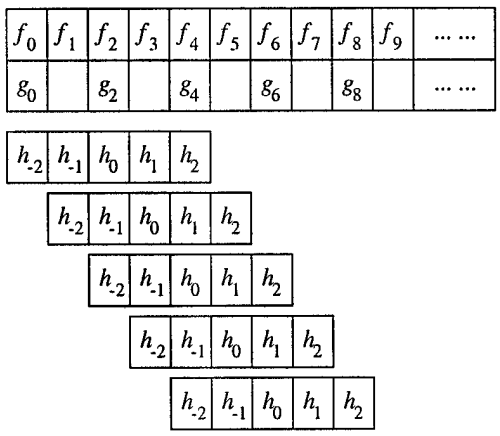


Figure 60. A 1-D example with every other sample missing in the measured signal.

We first write out all the terms involving a pixel (f_4) at which a measurement (g_4) was made in the noise term H_n according to Eq. (69) as Eq. (77),

$$\begin{aligned} E_4 &= ((f \otimes h)_2 - g_2)^2 + ((f \otimes h)_4 - g_4)^2 + ((f \otimes h)_6 - g_6)^2 \\ &= (f_0h_{-2} + f_1h_{-1} + f_2h_0 + f_3h_1 + f_4h_2 - g_2)^2 \\ &\quad + (f_2h_{-2} + f_3h_{-1} + f_4h_0 + f_5h_1 + f_6h_2 - g_4)^2 \\ &\quad + (f_4h_{-2} + f_5h_{-1} + f_6h_0 + f_7h_1 + f_8h_2 - g_6)^2 \end{aligned} \tag{77}$$

The derivative of H_n with respect to pixel f_4 can then be derived as Eq. (78), and further generalized as Eq. (79),

$$\frac{\partial H_n}{\partial f_4} = 2((f \otimes h)_2 - g_2)h_2 + 2((f \otimes h)_4 - g_4)h_0 + 2((f \otimes h)_6 - g_6)h_{-2} \tag{78}$$

$$\begin{aligned}\frac{\partial H_n}{\partial f_4} &= ((f \otimes h)_2 - g_2)h_2 + 0 \cdot h_1 + ((f \otimes h)_4 - g_4)h_0 + 0 \cdot h_{-1} \\ &\quad + ((f \otimes h)_6 - g_6)h_{-2} \\ &= (((f \otimes h) - g) \otimes h_{rev})_4\end{aligned} \tag{79}$$

where $h_{rev} = h_2\, h_1\, h_0\, h_{-1}\, h_{-2}$, and $(f \otimes h - g)$ is computed at all points where g_i is measured (that is, at g_2, g_4, and g_6).

We also write out all the terms involving an *unmeasured* pixel (say f_3) in the noise term H_n as Eq. (80).

$$\begin{aligned}E_3 &= ((f \otimes h)_2 - g_2)^2 + ((f \otimes h)_4 - g_4)^2 \\ &= (f_0h_{-2} + f_1h_{-1} + f_2h_0 + f_3h_1 + f_4h_2 - g_2)^2 \\ &\quad + (f_2h_{-2} + f_3h_{-1} + f_4h_0 + f_5h_1 + f_6h_2 - g_4)^2\end{aligned} \tag{80}$$

Again, the other three terms which *seem* to involve f_3 (convolutions centered at f_1, f_3, and f_5) are not in the summation. The derivative of H_n with respect to f_3 is derived in Eq. (81), and generalized in Eq. (82).

$$\frac{\partial H_n}{\partial f_3} = 2((f \otimes h)_2 - g_2)h_1 + 2((f \otimes h)_4 - g_4)h_{-1} \tag{81}$$

$$\begin{aligned}\frac{\partial H_n}{\partial f_3} &= 0 \cdot h_2 + ((f \otimes h)_2 - g_2)h_1 + 0 \cdot h_0 + ((f \otimes h)_4 - g_4)h_{-1} + 0 \cdot h_{-2} \\ &= (((f \otimes h) - g) \otimes h_{rev})_3\end{aligned} \tag{82}$$

Again, we compute the derivative at all points where g_i exists (that is, at g_2 and g_4).

We summarize the computation of the derivative of noise term with respect to a certain pixel in the following steps:

1) From f, compute $f_h = f \otimes h$;

2) Compute $n = \{n_i | n_i = \delta(i) \cdot (f_h - g)_i\}$, where $\delta(i)$ is 1 only at those points where g_i has a measurement, and 0 otherwise;

3) Compute $\frac{\partial H_n}{\partial f_i} = (n \otimes h_{rev})_i$.

Step 3 can be clarified with the illustration in Fig. 61, where n_i is non-zero only when i is even in this example.

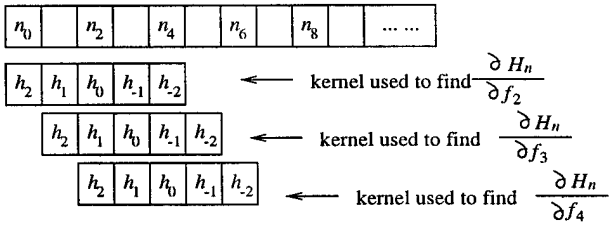


Figure 61. The reverse convolution in the derivation of the noise term.

We observe that the same kernel, h_{rev}, is actually used every time, but since some elements of n (those with odd indices) are zero, speed up may be accomplished by only using the elements of h_{rev} corresponding to even indices of n.

The convolution of 2-D images with the reverse kernel follows the same rule as that under 1-D case. Take the 2-D example that only one column of data (the fourth column) are missed, and the size of the blur kernel (h) is 3 x 3, as illustrated in Fig. 62. Similarly, f_{ij} represents a pixel from the original image, g_{ij} represents a pixel from the measured image, the question mark (?) represents unmeasured g_{ij} at that pixel, and h is the 3 x 3 blur kernel.

f_{00} g_{00}	f_{01} g_{01}	f_{02} g_{02}	f_{03} ?	f_{04} g_{04}	f_{05} g_{05}
f_{10} g_{10}	f_{11} g_{11}	f_{12} g_{12}	f_{13} ?	f_{14} g_{14}	f_{15} g_{15}
f_{20} g_{20}	f_{21} g_{21}	f_{22} g_{22}	f_{23} ?	f_{24} g_{24}	f_{25} g_{25}
f_{30} g_{30}	f_{31} g_{31}	f_{32} g_{32}	f_{33} ?	f_{34} g_{34}	f_{35} g_{35}
f_{40} g_{40}	f_{41} g_{41}	f_{42} g_{42}	f_{43} ?	f_{44} g_{44}	f_{45} g_{45}

$h_{-1,-1}$	$h_{-1,0}$	$h_{-1,1}$
$h_{0,-1}$	$h_{0,0}$	$h_{0,1}$
$h_{1,-1}$	$h_{1,0}$	$h_{1,1}$

Figure 62. A 2-D example with one column of data (the fourth column) missed.

Similar to the derivation steps in 1-D case, we first write out all the terms involving a pixel (f_{22}) at which a measurement (g_{22}) was made in the noise term H_n as Eq. (83),

$$\begin{aligned} E_{22} = {} & ((f \otimes h)_{11} - g_{11})^2 + ((f \otimes h)_{12} - g_{12})^2 + ((f \otimes h)_{21} - g_{21})^2 \\ & + ((f \otimes h)_{22} - g_{22})^2 + ((f \otimes h)_{31} - g_{31})^2 + ((f \otimes h)_{32} - g_{32})^2 \end{aligned} \tag{83}$$

where the convolution of f with the blur kernel h is extended as Eq. (84).

$$(f \otimes h)_{ij} = \sum_{m=-1}^{1} \sum_{n=-1}^{1} f_{i+m,\, j+n} \cdot h_{m,\, n} \tag{84}$$

The derivative of H_n with respect to pixel f_{22} can then be derived as Eq. (85), and further generalized as Eq. (86),

$$\begin{aligned} \frac{\partial H_n}{\partial f_{22}} = {} & 2((f \otimes h)_{11} - g_{11}) \cdot h_{1,\,1} + 2((f \otimes h)_{12} - g_{12}) \cdot h_{1,\,0} + 0 \cdot h_{1,\,-1} \\ & + 2((f \otimes h)_{21} - g_{21}) \cdot h_{0,\,1} + 2((f \otimes h)_{22} - g_{22}) \cdot h_{0,\,0} + 0 \cdot h_{0,\,-1} \\ & + 2((f \otimes h)_{31} - g_{31}) \cdot h_{-1,\,1} + 2((f \otimes h)_{32} - g_{32}) \cdot h_{-1,\,0} + 0 \cdot h_{-1,\,-1} \end{aligned} \tag{85}$$

$$\frac{\partial H_n}{\partial f_{22}} = (((f \otimes h) - g) \otimes h_{rev})_{22} \tag{86}$$

where h_{rev} has the form as indicated in Eq. (87), and $((f \otimes h) - g)$ is computed at all points where g_{ij} exists (that is, at g_{11}, g_{12}, g_{21}, g_{22}, g_{31}, and g_{32}).

$$h_{rev} = \begin{bmatrix} h_{1,\,1} & h_{1,\,0} & h_{1,\,-1} \\ h_{0,\,1} & h_{0,\,0} & h_{0,\,-1} \\ h_{-1,\,1} & h_{-1,\,0} & h_{-1,\,-1} \end{bmatrix} \tag{87}$$

We also write out all the terms involving an unmeasured pixel (say f_{23}) in the noise term H_n as Eq. (88).

$$E_{23} = ((f \otimes h)_{12} - g_{12})^2 + ((f \otimes h)_{14} - g_{14})^2 + ((f \otimes h)_{22} - g_{22})^2 + ((f \otimes h)_{24} - g_{24})^2 + ((f \otimes h)_{32} - g_{32})^2 + ((f \otimes h)_{34} - g_{34})^2 \quad (88)$$

The derivative of H_n with respect to f_{23} is derived in Eq. (89), and generalized in Eq. (90).

$$\frac{\partial H_n}{\partial f_{23}} = 2((f \otimes h)_{12} - g_{12}) \cdot h_{1,1} + 0 \cdot h_{1,0} + 2((f \otimes h)_{14} - g_{14}) \cdot h_{1,-1} + 2((f \otimes h)_{22} - g_{22}) \cdot h_{0,1} + 0 \cdot h_{0,0} + 2((f \otimes h)_{24} - g_{24}) \cdot h_{0,-1} + 2((f \otimes h)_{32} - g_{32}) \cdot h_{-1,1} + 0 \cdot h_{-1,0} + 2((f \otimes h)_{34} - g_{34}) \cdot h_{-1,-1} \quad (89)$$

$$\frac{\partial H_n}{\partial f_{23}} = (((f \otimes h) - g) \otimes h_{rev})_{23} \quad (90)$$

where h_{rev} has the same form as Eq. (87), and $((f \otimes h) - g)$ is computed at all points where g_{ij} exists (that is, at g_{12}, g_{14}, g_{22}, g_{24}, g_{32}, and g_{34}).

We summarize the computation of the derivative of noise term with respect to a certain pixel under 2-D case in the following steps:

1) From f, compute $f_h = f \otimes h$;

2) Compute $n = \{n_{ij} | n_{ij} = \delta(i, j) \cdot (f_h - g)_{ij}\}$, where $\delta(i, j)$ is 1 only at those points where g_{ij} has a measurement, and 0 otherwise;

3) Compute $\frac{\partial H_n}{\partial f_{ij}} = (n \otimes h_{rev})_{ij}$.

The following program implements this algorithm in 2-D.

```
for (r=0; r<rows; r++)
    for (c=0; c<cols; c++)
        fh[r][c] = convolve(f, h, r, c);
for (r=0; r<rows; r++)
    for (c=0; c<cols; c++)
        if (measured(r,c))
            n[r][c] = fh[r][c]-g[r][c];
        else
            n[r][c] = 0;
for (r=0; r<rows; r++)
    for (c=0; c<cols; c++)
        gradf[r][c] = convolve_speedup(n, hrev, r, c);
```

7.11 Complexity Analysis

Before comparing performance of the consistency methods and the MFA method using experimental techniques, we first analyze the algorithm complexity. We count each addition, subtraction, multiplication, division, or square root as one flop.

For the consistency method using NLSE criterion, if the dimension of image is m x n where m and n are large, and b is the number of brightness level, the restoration work is dominated by the operations of HHQR for the two linear systems and the consistency testing which takes b iterations for each row. The complexity of HHQR algorithm is on the order of $O\left(2mn^2 - \frac{2}{3}n^3\right)$ [Trefethen and Bau 97], and that of the consistency

testing is on the order of $O(2bmn^2)$. Therefore, the complexity of the consistency method is on the order of $O\left((2+2b)mn^2 - \frac{4}{3}n^3\right)$, where two HHQR is counted.

The complexity for the MFA method is not deterministic since the times of annealing is image-dependent. It takes $O((2pq+5)mn)$ flops for noise term related operations $O((2s+2t+2st+32)mn)$ flops for prior term related operations, and $O(5mn)$ for the update operations, where p x q is the dimension of blur kernel, s x t is the dimension of convolution kernel that simulates nth derivatives. Thus, the complexity of each annealing is on the order of $O((42+2pq+2s+2t+2st)mn)$. The number of annealing iterations is determined by the initial temperature, the final temperature, and the decreasing step of the temperature, which are chosen image-dependent. If the initial temperature is 10, final temperature 0.1, and decreasing step is 0.99, then the annealing times is 10000.

Therefore, generally, for image dimension on the order of 1000 x 1000, the MFA method has higher complexity than the consistency method. However, there is a straightforward hardware implementation of MFA restoration [Bilbro and Snyder 90][Bilbro et al. 98] which is very efficient.